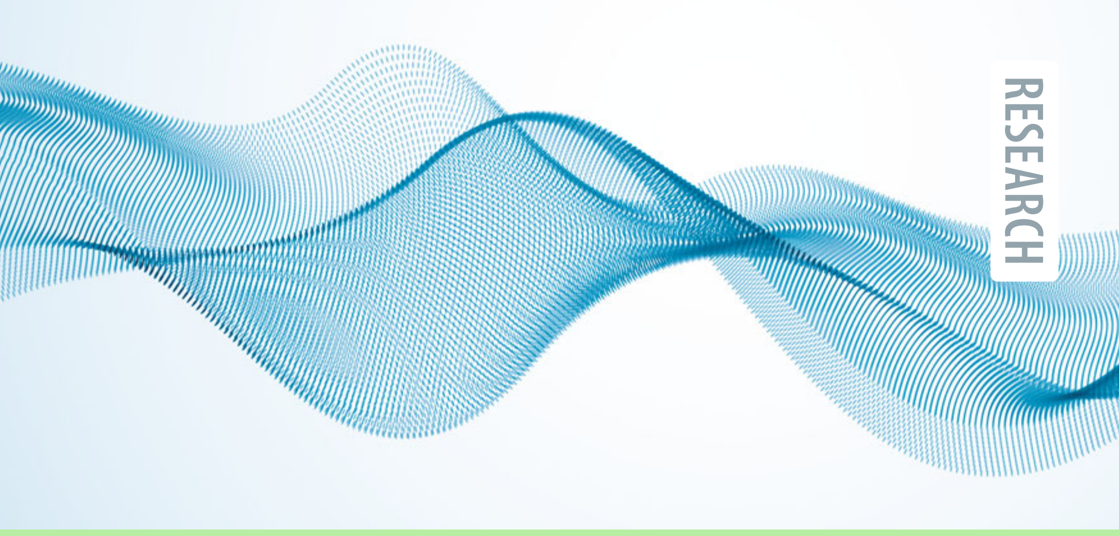


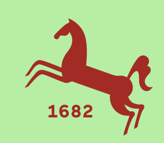
DIGITALE LITERATURWISSENSCHAFT

Mareike K. Schumacher

Orte und Räume im Roman

Ein Beitrag zur digitalen Literaturwissenschaft





J.B. METZLER

Digitale Literaturwissenschaft

Die Schriftenreihe ist ein Forum für literaturwissenschaftliche Arbeiten, die Forschungsfragen mit Hilfe digitaler Methoden zu lösen versuchen. Sie widmen sich literaturgeschichtlichen Themen oder Problemen der Literaturtheorie ebenso wie dem medialen Wandel oder kulturellen Kontexten. Charakteristisch ist die jeweilige fachliche Fundierung der digitalen Analysen, die in der Durchführung transparent sind und im Ergebnis so kommuniziert werden, dass ein breites Publikum damit weiterarbeiten kann. Die Reihe ist der nachhaltigen Fachwissenschaft verpflichtet und setzt neben eingängigen Argumentationen auch die Reproduzierbarkeit sämtlicher Datenanalysen voraus.

Mareike K. Schumacher

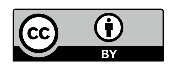
Orte und Räume im Roman

Ein Beitrag zur digitalen Literaturwissenschaft



Mareike K. Schumacher
Hamburg, Deutschland

Dissertation zur Erlangung des Grades der Doktorin/des Doktors der Philosophie an der Fakultät Geisteswissenschaften der Universität Hamburg im Promotionsfach neuere deutsche Sprache und Literatur vorgelegt von Mareike K. Schumacher Hamburg, 2021.



ISSN 2731-4022 ISSN 2731-4030 (electronic)
Digitale Literaturwissenschaft
ISBN 978-3-662-66034-8 ISBN 978-3-662-66035-5 (eBook)
https://doi.org/10.1007/978-3-662-66035-5

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Planung/Lektorat: Marta Schmidt
J.B. Metzler ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Aber falls also der Ort Existenz besitzt, so ist doch die Frage schwierig, was er denn nun ist, ob eine Art Körperausmaß oder eine andere Art von Wesenheit.

[Aristoteles, 1995]

Danksagung

Dieses Buch basiert auf meiner gleichnamigen Dissertation, die ich im Juni 2021 an der Universität Hamburg eingereicht und im Dezember desselben Jahres dort verteidigt habe. Neben meiner eigenen einige Jahre währenden Forschungsarbeit steckt darin die Hilfe vieler unsichtbarer Hände, die ich hier zeigen und denen ich von Herzen danken möchte. Allen voran gilt mein Dank meinem Doktorvater Jan Christoph Meister, der diese althergebrachte und inzwischen auch etwas umstrittene Bezeichnung, für mich im besten Sinne ausgefüllt und mit so viel positiver Bedeutung aufgeladen hat. Mit schier endlos scheinender Geduld und wohl überlegten Einwürfen – wie z. B. auch der Idee, mal nach einem Erzähltext zu suchen, der gar nicht über Raumdarstellungen verfügt – hat er den Entstehensprozess dieser Forschungsarbeit sanft geleitet. Ihm verdanke ich aber weit mehr als das Gelingen dieser Arbeit, denn sowohl zur Narratologie als auch zur digitalen Literaturwissenschaft bin ich durch seine Lehre gekommen. Die Begeisterung und die Forscherneugier, die er dadurch in mir geweckt hat, haben meinen Weg in die Literaturwissenschaft entscheidend geprägt. Und dafür bin ich extrem dankbar! Ich danke auch meiner Zweitbetreuerin Julia Nantke für den Austausch zur literarischen Raumdarstellung und natürlich auch für das ungemein zügige Begutachten der Dissertationsschrift.

Sehr sehr dankbar bin ich auch meiner Familie, Nicolas und Christian Schumacher, für ihre Geduld und vor allem dafür, mich in stressigen Phasen immer wieder von der Arbeit abzulenken, mich zum Lachen zu bringen und mich auf das Wesentliche, nämlich das Lebensglück, zu konzentrieren. Christian danke ich außerdem für unermüdliches und ungezähltes Korrekturlesen dieser Arbeit. Ohne den journalistisch geprägten Blick beim Gegenlesen und Überarbeiten hätte dieser Text sehr anders ausgesehen. Wenn es mir am Ende gelungen sein sollte,

die Leser*innen dieses Buches gut abzuholen und durch das Thema „Orte und Räume im Roman“ zu führen, so ist das ganz wesentlich ihm zu verdanken.

Für’s Korrekturlesen danke ich außerdem meiner Kollegin Marie Flüh von der Universität Hamburg. Ihr literaturwissenschaftliches Lektorat hat mich vor so manch unnötiger Schleife oder Wiederholung bewahrt („hinfort damit!“) und auch sie hat mich immer wieder und vor allem in den zehrendsten Phasen des Dissertationsprozesses zum Lachen gebracht, z. B. indem sie kleine Anekdoten der Literaturwissenschaft als Kommentare beim Gegenlesen hinterlassen hat („Funfact:...“).

Ich danke außerdem meinem Kollegen Michael Vauth von der Technischen Universität Darmstadt für’s Prüfen des narratologischen Kapitels 3 zur Vermessung narrativer Raumdarstellung. Die Diskussion über europäische und anglo-amerikanische Narratologie, literaturwissenschaftliche Erzähltheorie und narrative Turn, haben nicht nur das Kapitel bereichert, sondern mir auch viel Stoff zum Nach- und Weiterdenken gegeben.

Ganz maßgeblich für diese Arbeit sind die Visualisierungen des Modells und der Raumkategorien. Diese hat Jan-Erik Stange von der Freien Universität Berlin für mich entwickelt. Aber nicht nur sein Expertenblick auf die visuelle Darstellung hat die Arbeit an diesem Buch bereichert, sondern vor allem auch die konzeptuellen Anregungen und die neue Perspektive auf meine bisherigen Überlegungen. Vielen Dank für die vielen Denkanstöße, die dadurch entstanden sind!

Mit Diskussionen über Erzähltextanalyse und Computational Literary Studies haben aber noch weit mehr Menschen diese Arbeit geprägt. Vor allem der Austausch mit anderen Doktorand*innen im Koloquium und mit Kolleg*innen von nah und fern auf DHd-Konferenzen oder bei Summerschools wie z. B. den legendären dänischen Summer Courses in Narrative Studies (SINS) und der nicht weniger einschlägigen europäischen Sommeruniversität für Digital Humanities Leipzig (ESUDH). Bei diesen Treffen habe ich neben den Inhalten noch etwas ebenso Wichtiges gelernt, nämlich welches „meine“ Community ist und in welchem Forschungsfeld, d. h. vor allem mit welchen Forscher*innen, ich jetzt und in Zukunft arbeiten möchte. Durch alle hier Genannten und Gemeinten habe ich es nicht nur geschafft, diese Arbeit abzuschließen, sondern auch meinen Ort und meine Räume in der Forschungslandschaft zu finden!

Schließlich möchte ich noch meinen Eltern Ute und Manfred Höckendorff und Großeltern Irma und Hans-Jürgen Diercks und meiner Schwester Birthe Höckendorff und ihrer Familie danken für die Unterstützung und den Stolz und die Liebe, die sie mir entgegen gebracht haben. Damit meine ich weit mehr als die Phase der Dissertation, denn schon von klein auf haben sie mir das Gefühl gegeben, alles

erreichen zu können, was ich erreichen möchte. Dazu gehört auch ein fröhliches Cross-Over von PC-Schulungen der eigenen Firma, zu denen mein Vater mich und meine Schwester mitnahm, Tanz- und Ballettstunden zu denen meine Mutter mich anmeldete und unermüdlich hinbrachte und das Theater und die Theaterpädagogik für die meine Schwester durch ihre Begeisterung die Neugier in mir weckte. Ich glaube nicht, dass sie damals ahnen konnten, wo diese Zusammenführung von Kulturellem und Digitalem einmal hinführen würde und doch war gerade das für mich am Ende richtungsweisend. Ein Satz meines leider zu früh verstorbenen Großvaters hat die Arbeit an diesem Buch still voran getrieben. „Unterschätzt mir die Kleine nicht!" sagte er in meiner Kindheit bei jeder sich bietenden Gelegenheit zu anderen Erwachsenen, die in mir nicht mehr als das niedliche, blondgelockte, kleine Mädchen sahen, das ich damals gewesen bin. Ich bin sehr, sehr dankbar dafür, dass er in mir immer mehr gesehen hat als man von außen erkennen konnte und mich so, weit über seine Lebzeiten hinaus, darin bestärkt hat, meinen Weg zu gehen.

Inhaltsverzeichnis

Teil I

Das Modell



Einleitung

1

Betrachtungen zum Raum beschäftigen Forschende schon sehr lange. Bereits Aristoteles stellte sich die Frage, was Orte und Räume eigentlich sind. Was aber kann man heute einer bereits mehr als zweitausend Jahre währenden Forschungsarbeit unterschiedlicher Disziplinen noch hinzufügen? Das Forschungsfeld der Digital Humanities bietet eine Reihe von Methoden, die sich in besonderem Maße dazu eignen, alte Fragen auf neue Art und Weise zu betrachten, nämlich mit Hilfe von Computern und computationellen Verfahren (zur Computational Narratology (vgl. Mani, 2013, 84 ff) und zur computationellen Textanalyse (Gius und Jacke, 2015, passim)). Der Computer wird geradezu zum sokratischen Gegenüber – selber unwissend und darum immer wieder ein Spiegel für den Forschenden, der seine Fragen präziser formulieren, genauer fassen, immer weiter verfeinern muss, um schließlich im Laufe des Fragenstellens dem Erkenntnisinteresse ein Stückchen näher zu kommen. Zunächst einmal wissen Computer nichts über Orte und Räume und trotz der zahlreichen räumlichen Metaphern wie z. B. „das Netz", „der Cyberspace", „der Chatroom" oder „die Cloud" nehmen viele Aspekte des virtuellen Datentransfers keinen physischen Ort ein. Auf der einen Seite gibt es also eine Fülle von zum Teil jahrtausendealten Erkenntnissen, auf der anderen eine tabula rasa, die dabei helfen kann, diese auf neue Weise zu nutzen und auf einen klar definierten Forschungsgegenstand anzuwenden. Eine solche Verknüpfung theoretischer Traditionen und neuer informationstechnologischer Methoden zeigt auch diese Untersuchung. Dabei sind alle drei Parameter – Tradition, digitale Methoden und die Verknüpfung von beidem – in gleichem Maße von Bedeutung.

Seitdem Lessing im Jahre 1766 seine Laokoon-These aufgestellt hat, ist häufig der Gegenbeweis dafür angetreten worden. In zahlreichen Einzeltextanalysen, Analysen kleinerer Korpora oder theoretischen Betrachtungen wurde immer wieder dafür argumentiert, dass Rauminformationen für Literatur von ebenso großer Bedeutung sein können wie zeitliche Aspekte. In der Tat kann aus heutiger Sicht die

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_1

Annahme, dass Rauminformationen in Erzähltexten immer bedeutsam sind, als literaturwissenschaftliche Tatsache angenommen werden. Dennoch kann das Thema Raum in der Literatur nicht als abschließend erforscht und Lessings Laokoon-These weder eindeutig als verifiziert noch als falsifiziert betrachtet werden. Ein Grund dafür liegt darin, dass Lessing der Literatur nicht gänzlich abgesprochen hat, auch Referenzen auf den Raum zu enthalten (vgl. Lessing, 2012). Vielmehr stellte er fest, dass die Repräsentationsform von Raum in der Literatur eine andere sei als in der bildenden Kunst, was auch an der Linearität der textuellen Ausdrucksform läge. Poesie könne räumliche Aspekte immer nur aus einer Perspektive zeigen („Regel von der [...] Sparsamkeit in den Schilderungen körperlicher Gegenstände" (vgl. Lessing, 2012, 87)). Es kommen also qualitative und quantitative Aspekte zusammen; es mangelt der literarischen Raumdarstellung einerseits an Plastizität und andererseits werden räumliche Aspekte weniger ausführlich beschrieben als sie in der bildenden Kunst ausgedrückt werden. Bis heute ist noch keine Methode entwickelt worden, mit Hilfe derer der Referenzierung von Rauminformationen in Erzähltexten ein Wert zugewiesen werden kann, der einen Hinweis darauf gibt, wie groß der Anteil an Raumausdrücken und -hinweisen in einem Erzähltext tatsächlich ist. So konnte bisher auch noch nie eine komplexe Analyse sämtlicher Darstellungsweisen von Raum in der Literatur über ein vergleichsweise großes Korpus hinweg durchgeführt werden. Diese Lücke versucht diese Studie zu schließen. Verbunden mit der Frage nach dem Stellenwert von Raum in Erzähltexten wird betrachtet, ob sich der Anteil an Raumausdrücken und -beschreibungen im Laufe der Literaturgeschichte gewandelt hat. Zurückgreifend auf Lessings Laokoon-These aus dem Jahre 1766 und die vor allem nach dem spatial turn der Kultur- und Literaturwissenschaften zunehmenden Ansätze, die seine These falsifizieren (vgl. z. B. Genette, 1976, 43–48 und Sasse, 2009, 225) oder im Gegenteil auch verifizieren (vgl. Frank, 1977, passim), möchte ich die Arbeitshypothese aufstellen, dass zu Lessings Lebzeiten, also im 18. Jahrhundert, die Darstellung von Raum einen kleineren Anteil an Erzähltexten hatte als an zeitgenössischen und dass die Entwicklung zu stärker von Raumdarstellungen dominierten Erzähltexten diachron verlief. Darüber hinaus werden Muster von Rauminformationen in literarischen Texten betrachtet. Einzelne Wörter und Phrasen, in denen Informationen zum Raum enthalten sind, werden aus der Linearität der Texte extrahiert, um sie in ihrer Gesamtheit zu betrachten und so die grundsätzliche Beschaffenheit der Raumdarstellung in ihrer Mehrdimensionalität beschreiben zu können.

Gegenstand der Betrachtung ist ein Textkorpus, bestehend aus 100 Romanen des 18.–21. Jahrhunderts. Romane wurden als epische Langform unter der Vorannahme ausgewählt, dass sie nicht nur eine besondere Häufung von Informationen zum Raum beinhalten, sondern auch die größte Varianz in der Art der Darstel-

lung zeigen. Um einerseits einen Querschnitt durch die Literatur eines bestimmten Jahrhunderts zu erhalten, andererseits aber die Gesamtmenge der Texte in einem im Rahmen dieser Studie und mit den gewählten Tools betrachtbaren Umfang zu halten, wurde dieser auf 100 Texte festgelegt. Aus jedem der vier betrachteten Jahrhunderte werden also 25 Texte vergleichend im Hinblick auf den Stellenwert der Raumdarstellungen analysiert. Um dieses Textkorpus in einem angemessenen zeitlichen Rahmen betrachten zu können, wurden digitale Methoden eingesetzt, mit Hilfe derer die Raumausdrücke und -hinweise in den Erzähltexten ausfindig gemacht wurden.

„Welchen Stellenwert hat die Referenzierung von Rauminformationen in Romanen des 18.–21. Jahrhunderts?“, mit dieser Leitfrage möchte ich das dieser Untersuchung zu Grunde liegende Korpus systematisch und umfassend digital untersuchen. Diese Herangehensweise schließt bewusst einige Aspekte des Themas Raum in der Literatur aus. Da Raum hier ausschließlich auf Ebene des literarischen Textes betrachtet wird, ist weder die Perspektive des Schreibraums noch die des Leseraums von Interesse. Sowohl sehr stark autor*innenzentrierte als auch rezeptionsästhetische Ansätze werden also nicht betrachtet. Da aber das literarhistorisch uneinheitliche Korpus die Vermutung nahe legt, dass unterschiedliche Erkenntnisse zum Thema Raum bei der Textproduktion auch die textimmanenten Referenzen auf den Raum beeinflusst haben können, kann hier eine historische Perspektive nicht unberücksichtigt bleiben. Der wissenschaftsgeschichtliche Kontext, in dem ein Werk entstanden ist, findet also durchaus Berücksichtigung (und damit auch indirekt das Raumverständnis der Autor*innen).

Die Leitfrage impliziert aber auch, dass ein quantitativ erfassbarer Wert ermittelt werden kann, der den Anteil von Raumdarstellungen in Romanen messbar macht. Wie bereits erwähnt, wurde dazu in den Literaturwissenschaften bisher keine Methode entwickelt. Darum wurde hier anschließend an die Modellentwicklung ein Verfahren erdacht, um die unterschiedlichen Darstellungsweisen von Raum, die mit Hilfe des operationalisierbaren Modells kategorisiert werden können, im Hinblick darauf, wie explizit Raum referenziert wird, zu bewerten. Dem Ansatz liegt die Idee zu Grunde, dass nicht binär zwischen Raumausdruck und nicht-Raumausdruck unterschieden werden kann, sondern dass Wörter und Phrasen mehr oder weniger explizit Rauminformationen referenzieren können. Dieser Gedanke findet sich ebenfalls schon bei Lessing in der Feststellung, dass die Schilderung von Körpern in der Literatur zwar auch Hinweise auf den Raum gebe, aber nur andeutungsweise, nämlich durch Handlungen (vgl. Lessing, 2012, 87). Das angewendete Verfahren, die unterschiedlichen Kategorien von Raumdarstellungen quantitativ einzuschätzen, stützt sich auf die Grundlagen der „fuzzy set theory“. Diese wird im Bereich der spatial humanities, einer primär am Raum interessierten Forschungsrichtung

der digitalen Geisteswissenschaften, z. B. dazu genutzt, eine Landschaft nicht in klar voneinander abgrenzbare Regionen (wie z. B. Berg – nicht-Berg) einzuteilen. Stattdessen werden diskrete Einheiten entwickelt, die zwischen diesen Polen liegen und die darüber hinaus auch quantifiziert werden können (vgl. Bodenhamer et al., 2010, 55). Ein Beispiel für die Quantifizierung einer räumlichen Einheit wie der eines Berges, wäre, Regionen, die als definitiv-nicht-Berg kategorisiert werden können, den Wert 0 zuzuweisen, der Spitze des Berges den Wert 5 und dazwischen Abstufungen einzuführen (vgl. Bodenhamer et al., 2010, 55).

Mit Hilfe eines eigens dafür entwickelten CRF-Classifiers wurden Indikatoren für die narrative Darstellung von Raum in allen 100 Erzähltexten des Korpus automatisch annotiert. Das Korpus ist diachron gegliedert in vier Teilkorpora, die Texte aus dem 18., 19., 20. und 21. Jahrhundert beinhalten. Diese Aufspaltung ermöglicht es einerseits, den Classifier so zu trainieren, dass dieser unabhängig von Epochen- und Genrespezifika der Texte eine relativ hohe Erkennungsgenauigkeit erreicht und gleichzeitig mit Trainingsdaten zu arbeiten, die den sprachlichen Wandel nicht gänzlich außer Acht lassen. Die Aufteilung in jahrhundertspezifische Korpora ist nicht inhaltlich bedingt, sondern orientiert sich an statistischen Tests während des Machine-Learning-Trainings, die gezeigt haben, dass Trainingsdaten mit einem ähnlichen Entstehungskontext wie Testtexte einen deutlichen Performanz-Gewinn zeigen. Da die entwickelte Methodik die Technik des maschinellen Lernens beinhaltet, ist es von Vorteil, wenn Trainingstexte und untersuchte Texte sprachlich ähnlich sind (mehr dazu im Kapitel 6). Die 100 Erzähltexte wurden zufällig aus einem größeren Korpus von 400 Erzähltexten ausgewählt. Weitere 80 zufällig ausgewählte Texte aus ebendiesem Korpus wurden als (ebenfalls diachron unterteilte) Trainingsdaten zur Entwicklung der Methodik genutzt. Auf Basis der digital unterstützten Analyse wurde für jeden Roman im Korpus ein Raumindexwert (RIW) ermittelt, der einen Hinweis darauf gibt, wie hoch der Anteil von Raumdarstellungen im gesamten Erzähltext ist. Anhand dieser Werte können die Romane vergleichend analysiert und interpretiert werden.

Dieses Buch dokumentiert also eine Analyse auf der Grundlage interdisziplinärer Betrachtungen des Raumes, die einen Beitrag zum Forschungsfeld der digitalen Literaturwissenschaften leistet. Die Verknüpfung von Literaturwissenschaft (insbesondere Erzähltheorie) und Digital Humanities wurde etwa seit Beginn des Jahrtausends fruchtbar gemacht (vgl. z. B. Meister, 2008 und Mani, 2013) und kann damit noch als relativ neue Herangehensweise betrachtet werden. Es ist eines der Hauptanliegen dieser Studie, aufzuzeigen, dass diese Verbindung bidirektional genutzt werden kann. Auf der einen Seite werden Konzepte aus der literaturwissenschaftli-

chen Forschung so aufbereitet, dass sie maschinenlesbar[1] sind und zur Weiterentwicklung von Programmen aus dem Bereich der Digital Humanities genutzt werden können. Auf der anderen Seite führt das Training dieser Programme zur Schärfung der Konzepte. Der Erfolg dieser beidseitigen Weiterentwicklung hängt stark davon ab, dass stets beide Felder mitgedacht werden. Den Kern der Untersuchung macht ein Mixed-Methods-Ansatz aus, der hermeneutische und informationstechnologische Verfahren so eng miteinander verknüpft, dass diese kaum noch auseinander gehalten werden können.

Wegweiser für diesen Band

Dieses Buch ist in insgesamt drei Teile untergliedert. Im ersten Teil steht zunächst die Operationalisierung narrativer Raumdarstellung im Fokus. Ein Ziel der hier dokumentierten Studie war es, ein Modell zu entwickeln, das Indikatoren für narrative Darstellung von Raum an der Oberfläche von Texten definiert und in ein Kategoriensystem überführt, das zur Analyse literarischer Texte dienen kann. Grundsätzlich kommen in dieser Studie zwei unterschiedliche Modell-Begriffe zur Anwendung. Im ersten Teil wird der Begriff hauptsächlich für das literaturwissenschaftliche Raummodell genutzt (ein theoriegetriebenes Modell wie bei Heßbrüggen-Walter, 2017 erwähnt). Im zweiten Teil geht es dann vor allem um die Classifier-Modelle, die beim NER-Training entstehen (ein Datenmodell nach Jannidis, 2017, 100). Um die Modell-Begriffe klar voneinander abzugrenzen wird für letztere die Spezifizierung „CRF-Modell“ genutzt. Das theoriesbasierte Modell und das darin eingearbeitete Kategoriensystem für die Analyse narrativen Raumes wird in Abschnitt 2.1 vorgestellt. Anschließend wird die Herleitung des Modells begründet, die grundsätzlich aus zwei Perspektiven erfolgt: Der literaturwissenschaftlichen und der phänomenologischen. Bei der Modellentwicklung folge ich zunächst Traditionslinien der analogen, nicht-digitalen Forschung. Vom Gegenstand ausgehend wird zuerst die literaturwissenschaftliche Forschung daraufhin untersucht, ob sie operationalisierbare Aspekte narrativer Raumdarstellung beinhaltet, die die Definition von Indikatoren narrativer Raumdarstellung und damit auch das Modell unterstützen. Ausgehend von der Frage, ob Erzähltexte vielleicht sogar ganz ohne Darstellungen des Raumes auskommen können (Abschnitt 2.2), nähere ich mich der Bedeutung dieses Phänomens für die Literatur. Ausgangspunkt ist auch hier Lessings Laokoon-These:

[1] Der Begriff „lesbar“ ist in diesem Zusammenhang nicht im Sinne einer Verstehensleistung gemeint, wie sie bei sogenannten machine-reading-comprehension-Systemen angestrebt wird, die auf Basis natürlicher Sprache Fragen zu Texten beantworten sollen (vgl. z. B. Liu et al., 2019; Baradaran et al., 2020). Stattdessen ist das Ziel dieser Arbeit lediglich die automatische Erkennung und Klassifizierung von Wörtern, die Raum referenzieren.

> Wenn es wahr ist, daß die Malerei zu ihren Nachahmungen ganz andere Mittel, oder Zeichen gebrauchet, als die Poesie; jene nämlich Figuren und Farben in dem Raume, diese aber artikulierte Töne in der Zeit; wenn unstreitig die Zeichen ein bequemes Verhältnis zu dem Bezeichneten haben müssen: So können neben einander geordnete Zeichen, auch nur Gegenstände, die neben einander, oder deren Teile neben einander existieren, auf einander folgende Zeichen aber, auch nur Gegenstände ausdrücken, die auf einander, oder deren Teile auf einander folgen. (Lessing, 2012)

Lange Zeit war Lessings Standpunkt richtungsweisend, sodass es heute mehr literaturwissenschaftliche Arbeiten zur Thematik der Zeit als zu der des Raumes gibt. Seit der zweiten Hälfte des 20. Jahrhunderts, spätestens seit dem sogenannten spatial turn in den 1970ern (vgl. Günzel, 2009a, 10–11 und Dennerlein, 2009, 6), wurden aber zunehmend Ansätze zum Raum erarbeitet, von denen für diese Arbeit hauptsächlich kulturwissenschaftliche, narratologische und Digital-Humanities-Analysen relevant sind. Bei der Auswertung der literaturwissenschaftlichen Forschung wird stets die Frage mitberücksichtigt, inwiefern die gewonnenen Erkenntnisse so aufbereitet, d. h. operationalisiert[2], werden können, dass sie für die computergestützte Analyse zugänglich sind. Unter dieser Perspektive betrachte ich den symbolischen, kulturellen und metaphorischen Raumbegriff, Raumthemen, Raum als Strukturphänomen von Texten und die Konstruktion literarischen Raumes zwischen Text und Leser*in.

Es folgt in Abschnitt 2.3 ein interdisziplinärer Überblick zum Phänomen des Raumes, wie es jenseits der Literaturwissenschaften erforscht wurde. Ansätze aus den Kulturwissenschaften, der Soziologie, der Physik, der Mathematik, der Philosophie und der Archäologie werden so zusammengefasst, dass deutlich wird, welche unterschiedlichen Raum-Konzepte bestehen. In dieser Auswertung wird herausgearbeitet, wie z. B. der objekt-zentrierte, der relationale oder der soziale und kulturelle Raumbegriff im Gesamtdiskurs nebeneinander, einander überlagernd und zum Teil auch einander widersprechend bestehen. Das entwickelte Raummodell dient als Schlüssel für und Bindeglied zur digitalen Methodik, die im zweiten Teil dieser Studie einen Schwerpunkt bildet. An die theoriegeleiteten Betrachtungen anschließend, werden in Kapitel 3 die erzähltheoretischen Grundlagen dieser Arbeit dargelegt. Der Gegenstand dieser Arbeit wird in zwei Schritten definitorisch erfasst: Zuerst wird die Frage beantwortet, wie in dieser Studie der Begriff der Erzählung verstanden wird (Abschnitt 3.1) und dann wird noch einmal konkret zusammengefasst, was – dem theoriebasierten Modell folgend – die Referenzierung von Rauminformationen in Erzähltexten ausmacht (Abschnitt 3.2). Der erste Teil dieses Buches schließt in

[2] Operationalisierung wird hier im Sinne Morettis (Moretti, 2014) als ein Prozess verstanden, bei dem ein Konzept (hier literarischer Raum) in eine Reihe von Operationen zur Quantifizierung (hier z. B. Auffinden der räumlichen Ausdrücke in Romanen mit Hilfe von maschinellem Lernen) umgewandelt wird, die nötig sind, um es möglichst genau zu beschreiben.

Kapitel 4 mit den Annotationsguidelines für die im Modell integrierten sechs Kategorien literarischen Raumes und für Raummetaphern ab. Hier werden die Kategorien narrativer Raumindikatoren definiert, Annotationsspannen festgelegt und Beispiele gezeigt. Diese Annotationsguidelines bilden die Basis für das Machine-Learning-Training, das im zweiten Teil dieses Buches beschrieben wird.

Im zweiten Teil der Studie steht dann die digitale Methode im Vordergrund. Das Studiendesign folgt einem Mixed-Methods-Ansatz, bei dem eine Machine-Learning-Methode (Named Entity Recognition) im Zentrum steht und während der Analyse bedarfsorientiert durch weitere Methoden ergänzt wird. Die Grundlagen dieser Kernmethode, deren bisherige Anwendung in den Literaturwissenschaften und eine Methodenkritik werden in Kapitel 5 dargelegt. Im darauf folgenden Kapitel 6 wird der Machine-Learning-Prozess ausführlich dokumentiert. Dieser Prozess ist für diese Studie zentral, denn er wird keinesfalls als rein technisches Prozedere zur Optimierung einer Software verstanden, sondern, wie oben bereits erwähnt, als Kern eines Erkenntnisprozesses (zur Bedeutung informationstechnologischer Methoden für hermeneutische Forschungsprozesse vgl. Gius und Jacke, 2015). Das im ersten Teil der Arbeit entwickelte theoriebasierte Raummodell und der im Machine-Learning-Prozess für die automatische Annotation von Raumdarstellungen optimierte Classifier werden gemeinsam zur Basis einer in Kapitel 7 entwickelten Formel, mit Hilfe derer Rauminformationen in literarischen Erzähltexten quantifiziert werden können.

Den dritten Teil dieses Buches bildet eine analytische Anwendung der entwickelten Methode. Anhand einer Auswahl von 100 literarischen Erzähltexten des 18.–21. Jahrhunderts, die hier als Forschungsgegenstand dienen und darum als Kernkorpus bezeichnet werden, werden in Kapitel 8 beispielhaft Analysen auf drei Ebenen und aus drei Perspektiven durchgeführt. Auf Ebene des gesamten Kernkorpus (Abschnitt 8.1) werden unterschiedliche Aspekte literarischen Raumes, wie z. B. dessen grundsätzlich relationaler Aufbau, die Darstellung von Innen- und Außenwelten oder der diachrone Wandel einzelner Begriffe und deren Bedeutungszusammenhänge untersucht. In diesem Teilkapitel findet die Frage nach der Anschlussfähigkeit der Datenauswertung des digital unterstützen Ansatzes an traditionellere, nicht digitale literaturwissenschaftliche Forschung besondere Berücksichtigung. Im Abschnitt 8.2 wird auf Ebene von vier jahrhundertspezifischen Teilkorpora (je 25 Romane des 18.,19.,20. und 21. Jahrhunderts) eine explorative Untersuchung durchgeführt. Hier steht die Frage im Fokus, inwiefern auch eine textimmanente, datenzentrierte Betrachtung literaturwissenschaftlich relevante Einsichten hervorbringen kann. Abschließend zeigt Abschnitt 8.3 auf Ebene einzelner Textbeispiele, auf welche Weise Ausreißer-Phänomene, unabhängig von den in 8.1 und 8.2 gewählten Forschungsperspektiven, die Ergebnisse beeinträchtigen und sogar erheblich verzer-

ren können. Anschließend an diese Tradition und digitale Methodik verknüpfende Anwendung wird zum Abschluss in Kapitel 9 der Bogen zurück zum Ausgangspunkt geschlagen, der Frage nach der Quantifizierung narrativer Raumdarstellung. Mit Hilfe der in Kapitel 7 entwickelten Formel werden für jeden Text im Kernkorpus Indexwerte für die Referenzierung von Rauminformationen entwickelt. Der Vergleich dieser Werte zeigt die diachrone Entwicklung literarischer Raumdarstellung nicht anhand einzelner Phänomene, sondern in Form eines relationalen Wertes, der alle Phänomene (oder auch eine Gruppe von Phänomenen) literarischen Raumes zusammenfasst und so den literarhistorischen Vergleich ermöglicht.

Inwiefern die angestrebte Verknüpfung theoretischer Traditionen und digitaler Methoden geeignet ist, um zu neuen Antworten auf alte Fragen zu kommen und wie die hier erstellten Tools, Methoden und Daten nachgenutzt werden können, die Beantwortung dieser Fragen schließt diesen Band in Kapitel 10 ab. Die drei Teilaspekte von Forschungstradition, Innovation und Erkenntnisgewinn durch Verknüpfung derselben werden hier vor allem in ihrer Relevanz für die derzeitige literaturwissenschaftliche Forschung bewertet.



2 Modellierung narrativer Raumindikatoren

Zentral für die Betrachtung narrativer Raumdarstellung mit Methoden der Digital Humanities, genauer mit Machine Learning, ist ein Kategoriensystem, mit Hilfe dessen Indikatoren für Raumdarstellungen im Text identifiziert werden können. Ebenso zentral ist, dass dieses Kategoriensystem auf der einen Seite theoretisch fundiert ist und sich auf der anderen Seite zu einem Modell zusammenführen lässt. Das Modell – in diesem Falle eine schematische Darstellung – bringt die Kategorien in einen mehrdimensionalen Zusammenhang. Auf diese Weise hilft es dabei, die Kategorien nicht nur getrennt voneinander zu betrachten, sondern über das Zusammenwirken nachzudenken. Ein Zusammenwirken, das, wie schon erwähnt, in mehreren Dimensionen stattfinden kann. Ein solches Modell sollte nicht statisch, sondern dynamisch anpassbar sein, um auf unterschiedliche literaturwissenschaftliche Studien angewendet werden zu können.

2.1 Das fuzzy-set-Modell zur Betrachtung der narrativen Raumdarstellung

Das Kategoriensystem für Indikatoren narrativer Raumdarstellung umfasst insgesamt sechs Kategorien. Neben Indikatoren, die in einem buchstäblichen Sinne narrative Raumdarstellung referenzieren, wurden auch Metaphern berücksichtigt, die zwar eher als Störfaktor betrachtet wurden, als solche aber beim Modelltraining zunächst Berücksichtigung finden mussten. Ein Ausschluss aus dem Machine-Learning-Training hätte bedeutet, dass nicht hätte getestet werden können, wie groß dieser Störfaktor tatsächlich ist. Metaphern werden darum zwar in dieser Studie mitgedacht, in das Raummodell aber nicht integriert.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_2

Aus der aristotelischen Philosophie (mehr dazu in Abschnitt 2.3.1) stammt die Idee der Lageveränderung als Basis allen Raumempfindens. Diese beinhaltet auch, dass Raum über Positionen und den Wechsel derselben bestimmbar ist. Die Unterscheidung zwischen Ort und Raum wurde von Descartes übernommen (mehr darüber in Abschnitt 2.3.2. Orte werden demnach als Positionen oder Lagebestimmungen, Raum als mehrdimensionales Konzept definiert. Die im Modell dargestellten Kategorien sind unabhängig von einer Unterscheidung zwischen faktualen und fiktiven Räumen, da hier angenommen wird, dass die Erzählerfiktion, die für die betrachtete Textsorte grundlegend ist, erlaubt, dass neben mimetischen Darstellungen von Raum auch unvollständige oder gar inkonsistente storyworlds referenziert werden können (vgl. Nünning et al., 2004, 58). Die Modellierung orientiert sich einerseits an der fuzzy-set-theory in ihrer Anwendung durch die spatial humanities und greift auf der anderen Seite Aspekte der mathematischen Methode der principle component analysis (PCA) auf. Dabei werden potentiell unendlich viele Dimensionen eines Phänomens zu wenigen, meist zwei oder drei, voneinander unabhängigen, vektoriellen Komponenten (den principle components) heruntergerechnet, um umfangreiche Datensätze anschaulich zu machen (zu Grundlagen der PCA vgl. Jackson, 2003, zur Anwendung innerhalb der Digital Humanities vgl. z. B. Schöch und Steffen, 2014 oder Leufkens, 2020)[1]. Nun ist ein zentraler Vorteil eines von Machine-Learning-Methoden unterstützten Ansatzes, dass ein theoriebasiertes Modell auf konkrete Daten angewendet werden kann und dass die Schwierigkeiten des ersten Anwendungsversuches dazu genutzt werden können, das Modell noch einmal zu reflektieren. Die erste, konkreteste Kategorie im Raummodell ist die des Ortes. Die anderen Anwendungskategorien sind räumlich. Die sechs Kategorien referenzieren Rauminformationen im Text auf unterschiedliche Weise und können mit Hilfe der visuellen Metaphern in Abbildung 2.1 veranschaulicht werden.

Orte	Orte sind häufig sehr konkret und über einzelne Wörter im Text referenziert, sind aber phänomenologisch geprägt davon, in Zentren und Peripherien unterteilbar zu sein und nicht klar nach außen abgegrenzt (wie das Beispiel des Berges in 1 verdeutlicht).

[1] Die principle components analysis wird an dieser Stelle nicht tatsächlich durchgeführt, sondern lediglich als abstraktes Denkmodell verwendet, um die in den Abschnitten 2.3.1 bis 2.3.3 aufgefächerten Aspekte (wie z. B. Relationalität, Objektivität, kulturelle Bedeutung, soziale Funktion) zusammenzuführen.

Relationen Relationen werden ebenso explizit und anhand einzelner Wörter im Text referenziert. Sie wirken allerdings nur als Verbindung und / oder Zwischenraum zweier anderer Größen, die ihrerseits zwar häufig ebenfalls Informationen zum Raum referenzieren, die dies aber nicht zwingend tun müssen.

Relationale Verben Auch relationale Verben rufen über einzelne Wörter Informationen zum Raum auf. Die Kategorie umfasst sowohl Bewegungsverben als auch Verben der Wahrnehmung, die im Husserl'schen Sinne eine Subjekt-Objekt-Relation eröffnen (vgl. 2.3.2) und weitere Verben, die auf Relationen hindeuten wie z. B. „tragen", „küssen", „umarmen", „begleiten", „begegnen" und andere. Mit Ausdrücken, die in diese Kategorie fallen, wird ausgedrückt, dass Zwischenräume überbrückt und Relationen verändert werden. Die Verbindung zu anderen Größen im Text ist aber weniger eng als bei Relationen. Informationen darüber, inwiefern der relationale Raum sich durch diese Art der Verben verändert, können sich auch an weit entfernten Positionen im Text finden.

Raumthemen Zu Raumthemen tragen viele Wörter bei, die sich an sehr unterschiedlichen Textstellen finden können und dennoch zusammenwirken. Es ergibt sich ein unsichtbares Muster.

Raumbeschreibungen Raumbeschreibungen stehen in engem Zusammenhang mit einer anderen Größe, deren Räumlichkeit näher beschrieben wird. Das kann ein Teil des Settings und damit eine andere Rauminformation sein. Es kann sich aber z. B. auch um eine Figur handeln. Der räumliche Aspekt der Informationen, die diese Wörter referenzieren, muss nicht im Vordergrund stehen. Es kann also sein, dass die Indikatoren dieser Kategorie vage auf Raum hindeuten.

Raumhinweise Raumhinweise sind in Texten zahlreich und überall zu finden. Sie bilden aber keine feste Struktur, wie die Raumthemen, sondern wirken eher assoziativ zusammen. Die Rauminformationen, die sie enthalten, sind sehr implizit.

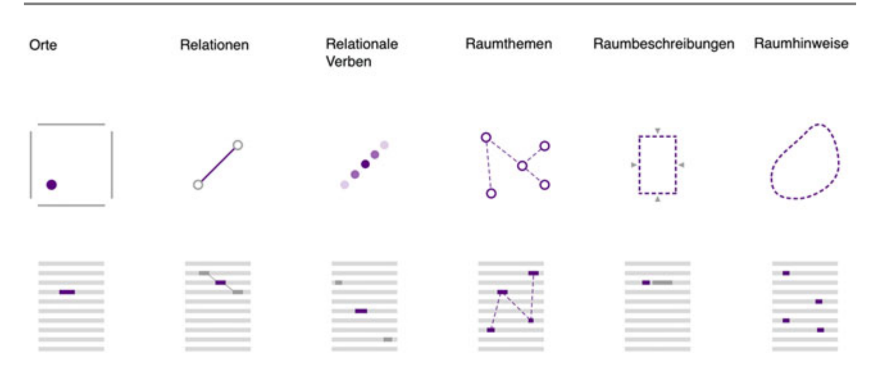


Abbildung 2.1 Kategoriensystem für Indikatoren narrativer Raumdarstellung

Insgesamt kann die Art der Referenzierung von Rauminformationen im Text auf einer Skala von explizit → implizit abgebildet werden. Auf dieser Skala können aber noch weitere Aspekte der Referenzierung räumlicher Informationen in Texten abgebildet werden, wie z. B. die Konkretheit des beschriebenen Phänomens oder auch, inwiefern Rauminformationen semiotisch im Vordergrund stehen. Der Vektor der Explizitheit wird darum, im Sinne der PCA, als erste zentrale Komponente literarischer Raumdarstellung definiert. Dieser erste Vektor bzw. diese zu einer Dimension zusammengerechnete Komponente beinhaltet Phänomene, die auf Ebene des Textes wirken.

Die sechs Kategorien können zusammengefasst werden zu einem fuzzy-set-Raummodell. Im Sinne der von (Zadeh, 1965) entwickelten Idee von fuzzy sets[2] gibt es Objektklassen, bei denen nicht klar, sondern nur graduell, über eine Zugehörigkeit entschieden werden kann. Solche Objektklassen sind zwar unpräzise definiert, aber fundamental für menschliche Kommunikation und Abstraktion (vgl. Zadeh, 1965, 338). Fuzzy ist das Raummodell einerseits, weil – je nach Raumbegriff – die einzelnen Kategorien mehr oder weniger zentral für ein Raummodell sind. Zwischen den Polen „definitiv zu einem Raummodell gehörig“ und „nicht Teil eines Raummodells“ gibt es graduelle Abstufungen, die die hier definierten Kategorien annehmen können. Eine solche Klassifizierung kann individuell durchgeführt und

[2] Dass dieser mathematische Ansatz auch in seiner Bedeutung für die Digital-Humanities-Forschung diskutiert wird, belegen z. B. Studien von (Martin-Rodilla et al., 2019) oder (Piotrowski, 2019) oder Konferenzen wie die, die unter dem Titel „Modelling Vagueness and Uncertainty in DH“ (2020) in Hamburg stattgefunden hat (Programm und Abstracts unter https://www.inf.uni-hamburg.de/inst/dmp/hercore/publications/vaguenessuncertainty2020.html [9.6.2021]).

das Raummodell somit als eine Art Werkzeugkasten benutzt werden, der für eigene Untersuchungen angepasst werden kann. Das Modell kann z. B. genutzt werden, um einen relationalen Raumbegriff zu betrachten, indem lediglich zwei der sechs Kategorien (Relationen und relationale Verben) in ein eigenes Raummodell übernommen werden. Je nachdem wie viele der Kategorien in ein eigenes Modell übernommen werden, ergibt sich ein engerer oder weiterer Raumbegriff. Werden z. B. nur Orte betrachtet, so entspricht dies einem sehr engen Raumbegriff, da die Kategorie „Orte“ als sehr nah beim Faktor „gehört definitiv zu einem Raummodell“ bewertet werden kann. Werden alle sechs Kategorien angewendet, so wird ein sehr weiter Raumbegriff betrachtet. In Enge und Weite des Raumbegriffs manifestieren sich mehrere andere Dimensionen von Raum. Physikalische Aspekte finden sich eher in den Kategorien enger Raumbegriffe, während soziale und kulturelle Dimensionen nur in weiten Raumbegriffen mit erfasst werden. Der Vektor eng → weit wurde aufgrund dieser Überlegungen als zweite zentrale Komponente (in Analogie zur PCA) ausgemacht. Dieser Vektor kann unabhängig vom Fokus auf Texte funktionieren. Beide Vektoren entbehren einen Nullpunkt. Die Nennung eines Ortes im Text z. B. ist maximal explizit im Sinne dessen, was an Explizitheit in einem semiotischen Modell möglich ist. Da immer auch implizite Informationen in einem Ortsausdruck enthalten sind, kann hier kein tatsächlicher Extremwert erreicht werden. Das Gleiche gilt für die Enge des Raumbegriffs. Ein hypothetischer maximal enger Raumbegriff müsste intersubjektiv als solcher definiert werden. Da aber jede Begriffsdeutung interpretationsabhängig ist, kann das, was aus einer Perspektive maximal eng wirkt aus einer anderen Teil einer weniger engen Konzeptionierung sein. Ein Nullpunkt ist auch hier nicht erreichbar.

In einem zweiten Sinne fuzzy ist das Modell durch die grundlegend nicht-binäre Beschaffenheit der einzelnen Kategorien, d. h. die zu einer Kategorie klassifizierten Ausdrücke können mehr oder weniger eindeutig zu dieser Kategorie gehören. Einem in einem Erzähltext namentlich referenzierten Fluss wie z. B. „Elbe“ kann ein höherer Wert in Bezug auf seine Zugehörigkeit zur Kategorie „Orte“ zugeschrieben werden als dem unspezifischen Ausdruck „Fluss“. Auch dieser Faktor ist interpretationsabhängig auch wenn im Rahmen dieser Studie die Interpretation darauf ausgerichtet ist, dass das Raummodell und die darin enthaltenen Kategorien anschlussfähig sind, indem die Kategorien so eindeutig wie möglich definiert werden. Dabei kommt erschwerend hinzu, dass die Kategorien des fuzzy-set-Raummodells nicht starr sind und einander überlappen können. Bei den im Kapitel 6 dokumentierten ausführlichen Tests des Raum-Classifiers hat sich herausgestellt, dass zum Teil Ambivalenzen zwischen den Kategorien bestehen. Andere sind klarer voneinander abgrenzbar. Generell gilt, dass es jeweils beliebig viele Wörter geben kann, die den Kategorien zugeordnet werden können und dass sie potentiell durchlässig oder auch

unterschiedlich stark ausgeprägt sind. Das zweidimensionale, skalar-vektorielle und auf sechs Kategorien beruhende fuzzy-set-Raummodell kann wie in Abbildung 2.2 visualisiert werden.

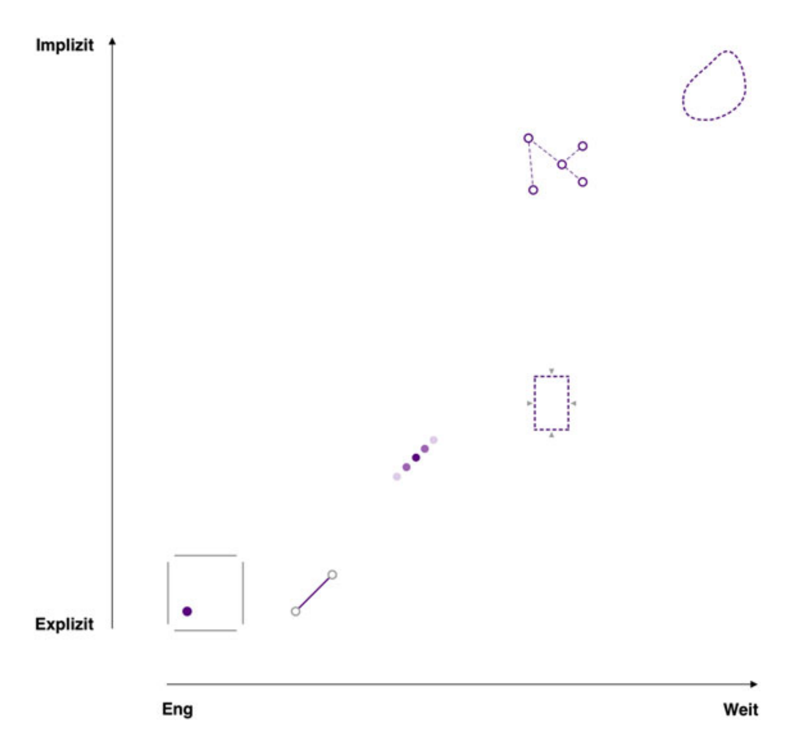


Abbildung 2.2 fuzzy-set-Raummodell

Die Inhaltskategorien der Raumausdrücke und -hinweise unterscheiden sich sowohl auf textlicher als auch auf semiotischer Ebene stark voneinander. Orte fungieren analog zu Punkten als kleinste Einheiten. Diese Punkte sind aber nicht hermetisch abgeschlossen, sondern können auch eine Durchlässigkeit an den Rändern aufweisen. Orte werden im Text ganz konkret benannt. Das gleiche gilt für Relationen. Dennoch referenzieren sie einerseits ein grundlegend anderes Phänomen und gehören auf der anderen Seite (zumindest potentiell) zu einem anderen Raumverständnis. Relationen sind Teil eines relationalen Raumbegriffs, der grundsätzlich etwas weiter geht als Raumkonzepte, die auf konkrete Ortsnennungen beschränkt sind. Relationale Verben referenzieren Rauminformationen viel impliziter als Orte

und Relationen. Sie können ganz andere Bedeutungen aufrufen, die sogar gegenüber den Rauminformationen im Vordergrund stehen können. Relationale Verben rufen häufig kulturell und historisch bedingte Assoziationsketten auf – wie z. B. „tanzen" in einem Erzähltext aus dem 18. Jahrhundert einen ganz anderen gesellschaftlichen Rahmen, ein ganz anderes Milieu impliziert als in einem Roman des 21. Jahrhunderts – und gehören darum zu einem erweiterten Raumbegriff. Auf ganz ähnliche Weise können Raumbeschreibungen lediglich implizit auf Informationen zum Raum verweisen. Sie geben zwar Hinweise darauf, wie ein Objekt im Hinblick auf seine Räumlichkeit beschaffen ist, ob es z. B. viel oder wenig Raum einnimmt, häufig ist das aber nicht die primäre Information. Auch Raumbeschreibungen gehören also zu einem erweiterten Raumbegriff. Einzelne Wörter, die zu Raumthemen beitragen, können dies auf sehr implizite Weise tun. Sie zeigen meist primär soziale, kulturelle und historische Aspekte. Das Wort „Kutsche" z. B. gehört zum Raumthema der Reise. Es deutet aber auch oder vor allem auf eine heute nicht mehr gebräuchliche, ziemlich beschwerliche Form der Reise hin, die vergleichsweise viel Zeit in Anspruch nimmt bzw. nahm. Raumthemen werden also nicht nur implizit referenziert, sie sind auch sehr stark in gesellschaftliche, historische und kulturelle Kontexte eingebettet und gehören darum zu einem weiten Raumbegriff. Raumhinweise referenzieren Raum schließlich lediglich indirekt und niemals als primäre Wortbedeutung. Die enthaltenen Rauminformationen wirken nur zusammen mit umfangreichem Kontext- bzw. Weltwissen. Wie viel Raum z. B. ein „Esstisch" einnimmt, der in einem Roman aus dem 19. Jahrhundert erwähnt wird, dessen Setting einen bürgerlichen Stadtteil von Paris beschreibt, ist nicht explizit, kann aber durchaus beim Lesen erfasst werden. Die maximal impliziten Raumhinweise, die eigentlich nur mit umfangreichem Kontextwissen dekodiert werden können, gehören zu einem sehr weiten Raumbegriff.

2.2 literaturwissenschaftliche Grundlagen der Raumdarstellung für das fuzzy-set-Modell

In den letzten Jahren sind einige umfassende Abhandlungen zur Thematik des Raumes in der Literatur erschienen, die sehr unterschiedliche Schwerpunkte bei der Betrachtung setzen. Im Jahr 2008 erschien Barbara Piattis literaturgeographische Abhandlung *Die Geografie der Literatur*. Hier werden in erster Linie die Relationen zwischen realweltlichen und in der Fiktion erwähnten Räumen untersucht. Piatti entwirft eine Reihe systematisch aufgebauter, abstrakter, grafischer Modelle, anhand derer das Verhältnis von kartographiertem Raum und Raum in der Literatur sehr genau bestimmt werden kann. In ihrem topologischen Ansatz berücksichtigt

sie vor allem konkret genannte Raumreferenzen auf Städte oder Landstriche, die auf Karten eingetragen werden können. Anhand eines kleinen Textkorpus zeigt sie mit Hilfe von Kartenvisualisierungen, welche Orte (Städte, Wege, Landschaften und Ähnliches) von besonderer Bedeutung für die betrachteten Texte sind. Da diese Kategorien sehr gut operationalisierbar sind und z. B. mit Hilfe von Listen und / oder Machine Learning sehr einfach automatisch in einem Text ausfindig gemacht werden können, ist dieser Ansatz auch für diese Studie von großer Bedeutung. Allerdings schließt Piatti mit ihrem Vorgehen viele erzähltheoretisch relevante Referenzen auf den Raum bewusst aus.

Dennerlein nähert sich dem Phänomen Raum in der Literatur hingegen von einer anderen Seite. In ihrer 2009 erschienenen *Raumnarratologie* unternimmt sie den Versuch, die bisherigen Betrachtungen der Raumkategorie in der Erzähltheorie zu systematisieren und zu einem eigenen Modell weiter zu entwickeln. Dabei konzentriert sie sich hauptsächlich auf textimmanente Merkmale, die den narrativen Raum konstruieren. Im Gegensatz zu Piatti betrachtet sie nicht nur Räume, die einen kartographierbaren Bezug aufweisen, sondern auch fiktive Räume oder Kategorien wie Innen- und Außenraum, räumliche Objekte und Raumreferenzen und -beschreibungen. Sie berücksichtigt nicht nur konkrete Raumreferenzen, sondern auch relationale Aspekte des Raumes und arbeitet gleichzeitig heraus, wie diese auf Wortebene ausgemacht werden können. Dabei schließt sie – ebenso wie Piatti (Piatti, 2008) vor ihr und Ryan et al. (Ryan et al., 2016) nach ihr – bewusst eine metaphorische Konnotation des Raumbegriffes aus. Im Hinblick auf die Operationalisierung des literarischen Raumes hat Dennerlein bereits wichtige Vorarbeit geleistet, auf die hier aufgebaut wird. Vor allem ihre Systematisierung räumlicher Bezüge in literarischen Texten, die eine konkrete Zuordnung sprachlicher Phänomene zu Kategorien von Raumreferenzen beinhaltet, ist für diese Studie von grundlegender Bedeutung. Eine Verbindung der beiden Ansätze der Literaturgeografie und der Raumnarratologie, für die Piatti auf der einen und Dennerlein auf der anderen Seite hier exemplarisch angeführt wurden, haben Ryan, Foote und Azaryahu in ihrem 2016 erschienenen Buch *Narrating Space/Spatializing Narrative. Where narrative Theory and Geography meet* hergestellt. Hier werden zunächst vor allem die Funktionen von Raum in Erzähltexten untersucht und dann die Rolle von Karten (materiellen und kognitiven), sowie Straßennamen und anderen geographischen Gegebenheiten, die ein Erzähltext widerspiegeln kann, betrachtet. Von den zwei Perspektiven, denen die Autor*innen hier nachgehen und die bereits im Titel benannt werden, ist für diese Studie nur eine essentiell. Die Idee, dass Raum durch Erzählen aufgebaut werden kann (narrating space), legt den Schluss nahe, dass dies mit bestimmten Bausteinen geschieht. Diese Bausteine narrativen Raumes in Form von konkreten Bezeichnungen räumlicher Elemente sind es, die für die Operatio-

nalisierung eines narrativen Raumkonzeptes als Vorarbeit zur Computerlesbarkeit herausgearbeitet werden müssen. Mit der Phrase *spatializing narrative* bezeichnen die Autor*innen einen Prozess, der zwischen Text und Leser*in bzw. interpretierender Person stattfindet. Der Erzähltext wird verräumlicht. Da in dieser Untersuchung aber nur der Text im Fokus steht, werden die Aspekte des *spatializing narrative* hier nicht berücksichtigt.

Auch seitens der kulturwissenschaftlich interessierten Literaturwissenschaft ist unlängst ein umfangreicher Sammelband mit Beiträgen zum Thema Raum in der Literatur erschienen (Dünne und Mahler, 2015). Hier wird zumeist ein anthropozentrisches Raumverständnis gegen einen eher mathematisch, geographisch oder topologisch geprägten Raumbegriff gestellt (vgl. Mahler, 2015, 17–19 und Nitsch, 2015, 31). Eine weitere Ebene, die hier den bisher beschriebenen Ansätzen hinzugefügt wird, ist, dass das Erzählen selbst als Kulturtechnik der Verräumlichung verstanden wird (Dünne, 2015, 44–45). Das Erzählen von Raum wird bei Dünne zu einem dynamischen konstitutiven Akt (Dünne, 2015, 44). Dies bedeutet, dass Raum nicht nur während des Erzählens für die Erzählung als solche entwickelt wird, sondern dass der Erzähltext als in sich geschlossenes Gesamtgefüge sich auch auf außertextuelle, kulturelle Räume auswirkt und diese prägt (Dünne, 2015, 49–52). Dieser Ansatz ist zwar an sich interessant und geht über das hinaus, was in anderen Traditionslinien der Betrachtung narrativen Raumes festgestellt wurde, in der vorliegenden Untersuchung kann genau dieser Aspekt allerdings keine Berücksichtigung finden, da er nicht unabhängig von einer Erzählsituation gedacht werden kann. Die Verräumlichung geschieht entweder beim Erzählen – bzw. im Falle der Literatur beim Schreiben – bzw., beim Zuhören oder beim Lesen. Beides lässt sich nicht in die hier gewählten computergestützten Verfahren integrieren.

Dass in den letzten Jahren nicht nur konzeptionelle Methoden, sondern auch konkrete informationstechnologische Verfahren aus dem Bereich der Geographie vermehrt in der gesamten geisteswissenschaftlichen Forschung eingesetzt wurden, zeigen Publikationen wie *Spatial Humanities* von (Bodenhamer et al., 2010), *Towards Spatial Humanities* von (Gregory and Geddes, 2014) oder *Hypercities. Thick Mapping in the Digital Humanities* von (Presner, 2014). Hier steht der Einsatz von Geoinformationssystemen (GIS) z. B. in den Geschichts- und Politikwissenschaften im Zentrum des Interesses. Dass Ansätze zur GIS-Technologie nicht nur für die geisteswissenschaftliche Forschung fruchtbar gemacht werden konnten, sondern dass andersherum auch narratologische Erkenntnisse in Theorien zur digitalen Kartographie einfließen, zeigen die Autor*innen von *Deep Maps and Spatial Narratives*, das 2015 von Bodenhammer, Corrigan und Harris herausgegeben wurde. Jenseits von geographischen Aspekten und Geoinformationssystemen als deren Visualisierungstools, gibt es aber auch Ansätze für eine digitale Raumnarratologie (Barth und

Viehhauser, 2017; Viehhauser, 2020). Hier wurde bereits begonnen, auch andere Informationstechnologien aus den Digital Humanities bei der Untersuchung narrativer Räume einzubeziehen. Dazu gehören z. B. der Einsatz von Tools aus dem Methodenkomplex des Natural Language Processing (NLP) und (manuelle) digitale Annotation. Auch Visualisierungen wie z. B. Netzwerkanalysen und Distributionsgraphen wurden hier bereits erprobt. An diese aktuellen Entwicklungen – die Verbindung von Raumgeographie und Raumnarratologie und deren Weiterentwicklung in den Digital Humanities – schließt die hier vorliegende Untersuchung an.

Raum wurde in der Literaturwissenschaft aber nicht nur unter geographischen, erzähltheoretischen, kulturwissenschaftlichen und informationstechnologischen Gesichtspunkten betrachtet. In unterschiedlichen, einander teilweise abwechselnden oder ergänzenden literaturwissenschaftlichen Traditionslinien wurde literarischer Raum unter sehr unterschiedlichen Gesichtspunkten untersucht. In Abschnitt 2.2.1 wird nachgezeichnet, inwiefern in den Literaturwissenschaften ein symbolischer, kultureller und metaphorischer Raumbegriff von Bedeutung ist. Im Anschluss stehen Raumthemen im Fokus, die in der literaturwissenschaftlichen Forschung besonders häufig betrachtete wurden (2.2.2). Eine weitere Betrachtungsweise, die in Abschnitt 2.2.3 beschrieben wird, ist die des literarischen Raumes als Strukturphänomen von Texten. Eine Herangehensweise an literaturwissenschaftlich relevante Raumphänomene, die für diese Studie weniger zentral ist, und in Abschnitt 2.2.4 darum nur kurz skizziert wird, ist es, Raum als Konstruktion zwischen Text und Leser*in zu sehen. Alle Aspekte der literaturwissenschaftlichen Analysen narrativen Raumes, die operationalisierbar sind und darum in die Modellentwicklung einbezogen wurden, werden in Abschnitt 2.2.5 zusammengefasst.

Neben diesen generellen Untersuchungen zum Raum in der Literatur wurden zahlreiche Einzeltextanalysen und Betrachtungen einzelner Phänomene in kleineren Korpora vorgelegt (vgl. z. B. Menzel, 2005; Gromes, 2005; Wintgens, 2005; Kurzenberger, 2005; Schneider, 2005; Whybra, 2005; Zetzmann, 2020; Rocchi, 2020 oder Hafner, 2020). Diese Einzeltextanalysen oder allzu stark auf kleine, homogene Korpora ausgerichtete Studien werden aber nicht zur Modellentwicklung herangezogen, da es ihnen zumeist an Übertragbarkeit mangelt. Stattdessen stehen Konzepte im Vordergrund, die zu einer allgemeinen Schärfung des Raumbegriffes in der Literatur beitragen. Die unterschiedlichen Ansätze werden anhand der Frage geprüft, ob sie zu einer Operationalisierung des Raumkonzeptes beitragen können, die zum Ziel hat, Raum in Erzähltexten computerlesbar zu machen. Dabei ist mit Lesbarkeit kein Textverständnis gemeint, sondern die Fähigkeit des Computers, Wörter zu erkennen (Erkennung), die zu Raumdarstellungen beitragen, und diese anhand vorher erlernter Kategorien zu annotieren (Kategorisierung).

2.2.1 Symbolischer, kultureller und metaphorischer Raumbegriff

Einer der frühen Ansätze zur Thematik des Raumes in der Literatur ist die Essay-Reihe *Spatial Form in Modern Literature* von (Frank, 1945a). In seinen Essays stellt Frank fest, dass die avantgardistische Literatur der Moderne über eine sehr spezifische Räumlichkeit verfügt (besonders deutlich weist Frank, 1977, 231 auf die enge Verknüpfung von Räumlichkeit und literarischer Moderne in seinem Essay *Spatial Form: An Answer to Critics* hin). Sowohl die Poesie als auch die Erzählliteratur entwerfen ein System aus Referenzen, welches sich dem Lesenden in einer Komplexität darstellt, die nur als räumlich und retrospektiv wahrgenommen werden kann. Diese retrospektive Erfassung der Räumlichkeit folgt einem Prinzip, das Frank als „principle of reflexive reference" (Frank, 1945b, 230) bezeichnet. In Franks Essays wird der Raumbegriff rein metaphorisch verwendet und steht im Grunde für ein System thematischer Referenzen, aus denen sich eine Struktur ergibt (Herman, 2010, 555). Anders als bei Rauminformationen generell kann hier zwischen Texten unterschieden werden, die diese Art der „spatial form" aufweisen und denen, die dies nicht tun (ebd.). Obwohl seine Feststellung, dass Texte Räumlichkeit besäßen also durchaus als bahnbrechend bezeichnet werden kann (auch wenn Frank (1977) selber sie eher als eine Ausnahme zur Bestätigung der Laokoon-These sieht denn als Falsifikation derselben), ist seine Theorie nur ansatzweise auf Texte übertragbar, die nicht zur Moderne gehören. Da sein Raumbegriff ein Phänomen bezeichnet, das zwischen Text und Leser*in entsteht, und das im Prinzip einem System aus nicht expliziten Relationen zwischen thematischen Einheiten von Texten entspricht, kann Franks Ansatz nichts zu einem Modell beitragen, das Raum in Erzähltexten computerlesbar macht.

Die Traditionslinie, die einen hauptsächlich thematisch-inhaltlichen und damit stärker textgebundenen Raumbegriff zum Gegenstand hat, beginnt bei Bachelard. In seiner *Poetik des Raumes* (1960) entwickelt er eine stark von der Psychoanalyse geprägte Typologie räumlicher Motive. Die Verknüpfung von Psychoanalyse und Poesie begründet er damit, dass fiktive Räume immer von einer Einbildungskraft geschaffen werden, die an Individuen gebunden ist (Bachelard, 1987, 30). Der reale Raum könne vermessen und objektiviert werden, der poetische Raum sei hingegen immer ein erlebter Raum (ebd.). Bachelard schlägt darum eine Topo-Analyse vor, die er als „systematisches, psychologisches Studium der Örtlichkeiten unseres inneren Lebens" beschreibt (Bachelard, 1987, 40). Bachelard entwickelt eine Raumsymbolik, die auf Dualismen fußt. Am Beispiel des Hauses verdeutlicht er, dass die vertikale Achse von oben nach unten eine Aufladung von positiv (oben) zu negativ (unten) erfährt. Darüber hinaus symbolisieren die oberen Räume wie z. B. der

Dachboden über ihre Nähe zum Himmel das Licht, die Aufklärung und insgesamt den Intellekt. Die Kellerräume stehen hingegen für Dunkelheit, Angst und Bedrohung. Verbunden sind diese Räume durch Treppen, die den Aufstieg, und damit das zunehmend Positive oder den Abstieg, und damit das zunehmend Negative, symbolisieren (Bachelard, 1987, 49–62). Neben dieser vertikalen Achse gibt es nach Bachelard die ebenfalls symbolisch zu deutende Zentralisierung von außen nach innen (Bachelard, 1987, 50). Auch für Klotz ist der Verweischarakter literarischen Raumes grundlegend (Klotz, 1969, 29). In seinen Einzelanalysen von Romanen aus unterschiedlichen räumlichen, sprachlichen und zeitlichen Zusammenhängen wird deutlich, dass Raum im Roman auf unterschiedlichen Ebenen eingesetzt werden kann. Auf Ebene der Gesamthandlung, auf Ebene einzelner narrativer Einheiten wie z. B. eines Dialogs und auf Wortebene (Klotz, 1969, 26). Als räumlich determinierte Mikro-Einheiten des Romans benennt er „Panorama“, „Tableau“ und „Stadt“. Darüber hinaus macht er im Verlauf seiner Betrachtungen einige wiederkehrende Raumsymbole aus, die für die Literatur von besonderer Bedeutung sind, wie z. B. das Fenster als Verbindung von Innen- und Außenwelt (Klotz, 1969, 176 und 349). Anders als Bachelard, der ein System fester Codes entwirft, betont Klotz stets die Dynamik räumlicher Codierung. War in einem Textbeispiel das Private noch das Schutz spendende Element, welches von der Stadt bedroht wurde, so kann diese Bedeutung in einem anderen Beispiel ins Gegenteil verkehrt sein und die Stadt wird zum Symbol der Geborgenheit, während das Private eine Bedrohung darstellt (Klotz, 1969, 283). Klotz erkennt in seinen Analysen außerdem, dass Raum bzw. in seinen Beispielen die Stadt und das Städtische für sich stehen, auf Figuren einwirken, zur Charakterisierung von Figuren beitragen oder sogar zum handelnden Individuum stilisiert werden kann (Klotz, 1969, 64, 152, 226 und 279). Ähnlich wie Bachelard verfolgt auch Lotman einen dualistischen Ansatz und entwickelt daraus ebenfalls ein Raummodell, das auf Achsen beruht. Ausgangspunkt ist der menschliche Körper als Wahrnehmungsinstanz von Raum. Wie bei Bachelard gibt es eine Achse von oben nach unten. Dazu kommt eine horizontale rechts-links-Achse, die eine ebenso positiv-negative Aufladung hat wie die vertikale Achse (Lotman, 1973, 313–326). Das Herausragende an Lotmans Ansatz ist die enge Verknüpfung seiner Raumsemantik mit kulturellen Modellen oder „Weltbildern“ (Lotman, 1973, 313–326). Statt, wie Bachelard, ein starres Deutungsmuster aufzubauen, ist Lotmans Ansatz – ähnlich wie der von Klotz – die Basis für ein dynamisches System, das offen für Umdeutungen ist. Er betont in erster Linie die Abhängigkeit der Belegung von Raumausdrücken mit nicht-räumlicher Bedeutung von den kulturellen Zusammenhängen, in denen diese stattfinden. Es handelt sich hierbei um kulturelle und damit vor allem auch kulturraumspezifische Codierungen (Lotman, 1973, 313–326). Diese kulturelle Codierung von Raum wird auch in späteren literaturwis-

senschaftlichen Arbeiten zum Raum immer wieder betont (z. B. von Bal and van Boheemen-Saaf, 1985, 138). Zusätzlich zu diesem an Achsen und Dualismen orientierten Ansatz führt Lotman die Grenze als zentrales literarisches Raumsymbol ein (Lotman, 1973, 328–329).

Der in der Literaturwissenschaft relativ häufig rezipierte Humangeograph Tuan hat in *Space and Place* (1977) eine Theorie entwickelt, die den Ideen Lotmans zur literarischen Raumdarstellung stark ähnelt. Auch Tuan setzt die Erfahrung des Menschen als körperliches Wesen an den Ausgangspunkt jeglicher Raumerkenntnis. Genauso spielen bei Tuan am Körper orientierte Achsen und die kulturelle Codierung ihrer Pole eine große Rolle. Statt jedoch wie Lotman die semantische Dimension zu betonen und den Verweischarakter herauszuarbeiten, hebt Tuan eher auf eine Metaphorik des Raumes ab (Tuan, 1979, 34–50). Ausgehend von der Erfahrung der meisten Menschen als Rechtshänder z. B. (Tuan, 1979, 43), entwickeln sich in ihrem Ursprung raummetaphorische Ausdrücke wie „der rechte Weg", „rechtschaffen" oder auch „er ist meine rechte Hand". Der Wortstamm „recht" hat in den unterschiedlichen Ausdrücken graduell von seiner Raumsemantik verloren und steht zum Teil (z. B. bei rechtschaffen) nur noch für den positiven Wert, der durch die menschliche Erfahrung der rechten Hand als derjenigen, mit der alles gut erledigt werden kann, hervorgerufen wird. Zurückgehend auf die Nähe der Ansätze von Lotman und Tuan wird in der heutigen Literaturwissenschaft meist verallgemeinernd von einem metaphorischen Raumbegriff gesprochen, um einen Rückbezug zu dieser eigentlich semantisch-kulturell-metaphorisch geprägten Tradition herzustellen (so z. B. bei Ryan et al., 2016, 17).

Die literaturwissenschaftlichen Ansätze von Bachelard, Klotz und Lotman werden von Hofmann in seiner Abhandlung *Raum, Situation, erzählte Wirklichkeit* (1978) weiter verfolgt und ausgebaut. Er arbeitet sehr deutlich heraus, dass die räumlichen Achsen, die sowohl bei Bachelard als auch bei Lotman die Basis der semantischen Verknüpfung bilden, von der Raumerfahrung des Menschen ausgehen (Hoffmann, 1978, 3–6). Dualismen wie oben-unten, nah-fern oder links-rechts gehen immer von einem Subjekt aus. Hinzu kommt, dass die Verknüpfung mit Qualitäten wie gut-schlecht, vertraut-fremd u.ä. auch immer von einem Subjekt (und nicht zwingend von einer ganzen Kultur wie bei Lotman) ausgehen muss (Hoffmann, 1978, 3–6). Um seinen subjektzentrierten Ansatz zu systematisieren, führt Hofmann auf unterschiedlichen Ebenen Unterscheidungskategorien ein, die jedoch weit davon entfernt sind, sich in ein Gesamtkonzept narrativen Raumes einfügen zu lassen. Seine Kategorien wie die Unterscheidung von gelebtem Raum und erzähltem Raum (Hoffmann, 1978, 1–3), Anschauungsraum, gestimmtem Raum und Aktionsraum (Hoffmann, 1978, 3–6) oder auch die zum Teil von Klotz übernommene Verwendung von Szene, Panorama und Tableau als räumliche Einheiten von Erzähltexten erweisen sich in seinen Einzeltextanalysen als unscharf. Die größte Stärke

von Hofmanns Ansatz liegt darum nicht bei der Entwicklung allgemein gültiger Analysekategorien, sondern in der engen Verknüpfung von Raum und Subjekt[3]. Damit zeigt er, was in Lotmans Modell bereits in Ansätzen und bei Klotz schon deutlich angelegt ist, nämlich, dass Raum dynamisch interpretierbar ist (Hoffmann, 1978, 9–11). Unterschiedliche Figuren können nicht nur im Erzähltext unterschiedliche Raumsemantiken nutzen, sondern die Codierung räumlicher Begriffe kann sich für eine Figur auch innerhalb der Erzählung wandeln. Außerdem schlägt Hofmann eine Typologie der Raumsymbolik vor, deren Kategorien sich vor allem durch Offenheit der Codierung voneinander abgrenzen (Hoffmann, 1978, 319–327). Hier integriert er binäre symbolische Raumrelationen wie oben-unten und symbolische Raumeinheiten wie das Haus oder die Grenze. Außerdem ergänzt er das kontextbezogene Raumsymbol und die Unterscheidung zwischen übergreifender Raumsymbolik eines ganzen Romans und einzelnen Raumsymbolen, die als narrative Einheiten wirken (Hoffmann, 1978, 319–327).

Von den bisher betrachteten Ansätzen dieser Traditionslinie sind vor allem zwei Einsichten für diese Studie von zentralem Interesse:

1. Raumreferenzen können mit einem Bewertungssystem verknüpft sein, welches auf eine Skala zurückgeführt werden kann, die von negativ bis positiv verläuft und
2. Raumausdrücke können ihre ursprüngliche, räumliche Bedeutung verlieren und nur noch metaphorisch gebraucht werden.

Es sollte also klar zwischen Räumen, die eine Bewertung erfahren, wie z. B. das Haus als Ort der Geborgenheit, und Raumausdrücken, die für einen Wert stehen, wie z. B. „er war am Boden zerstört“, unterschieden werden. In der Operationalisierung narrativer Raumdarstellung wurde außerdem berücksichtigt, dass semantisch aufgeladene Raumausdrücke vom Computer erkannt werden sollen, Raummetaphern hingegen nicht. Denn Raummetaphern beschreiben keine Orte und Räume, obwohl sie Ausdrücke enthalten, die üblicherweise genau dazu verwendet werden, sondern stehen immer für etwas anderes.

Ein anderes Oppositionspaar als das positiver und negativer Bewertung ist zentral für die feministische Narratologie, die an die oben skizzierte Traditionslinie anknüpft. Sowohl Bal (Bal and van Boheemen-Saaf, 1985, 137) als auch Nünning und Nünning (2004) stellen die Semantisierung des Raumes in Zusammenhang mit geschlechterspezifischen Strukturen. Wie bereits Klotz und Hofmann betonen auch

[3] Auf diese Subjektivität greift später z. B. auch Mieke Bal (Bal and van Boheemen-Saaf, 1985, 136) zurück.

Nünning und Nünning die enge Verbindung zwischen Räumen und Figuren. Unterschiedliche Figuren erleben unterschiedliche Räume auf ihre eigene Weise (Nünning et al., 2004), und Raumerfahrungen tragen entscheidend zur Identitätsbildung bei (Nünning et al., 2004, 56). Besonders bedeutsam für die gendertheoretische Betrachtung literarischen Raumes ist die Möglichkeit der Figuren, sich von einem zu einem anderen Ort zu bewegen und Grenzen zu überschreiten. Die Möglichkeit hierzu sei stets von den Machtverhältnissen geprägt (Nünning et al., 2004, 52). Diese literaturwissenschaftliche Forschungsrichtung berücksichtigt noch stärker als Vorgänger wie Lotman den sozialen Kontext von Erzähltexten. Wie Warhol (2012) am Beispiel von Jane Austen zeigt, kann das Wissen darüber, welche Räume zur Entstehungszeit des Textes mit welchen Genderstereotypen verbunden waren, die Interpretation leiten. Öffentliche Räume bzw. Räume, in denen Frauen zur Zeit Jane Austens wenig oder gar nicht präsent waren, spielen auch in ihren Texten eine untergeordnete Rolle (Warhol, 2012, 93). Das heißt auch, dass aus dieser Perspektive nicht nur erwähnte Räume oder Raumkategorien für Erzähltexte von Bedeutung sein können, sondern dass auch von der Erzählung weitgehend ausgeschlossene Räume im Hinblick auf eine feministische Interpretation zentral sein können. Zwar können mittels computergestützter Methoden in einem Text hauptsächlich Räume berücksichtigt werden, die auch erwähnt werden, doch bietet eine diachron angelegte Studie wie diese, die Möglichkeit, Differenzialanalysen durchzuführen. Für bestimmte Raumausdrücke, Kategorien oder Gruppen von Raumausdrücken und -kategorien kann in einem Korpus eine Art „Normalverhalten“ ermittelt werden. Diese kann dann mit einem kleineren Textkorpus oder auch einem Einzeltext verglichen werden. Eine solche Differenzialanalyse wird in Abschnitt 8.1.4 anhand des Weltbegriffs bei Wilhelm Raabe beispielhaft durchgeführt. Außerdem kann dieser Ansatz fruchtbar gemacht werden, indem Raumdarstellungen mit anderen narratologisch bedeutsamen Größen wie z. B. Figuren zusammen betrachtet werden. Zwar scheint die Verknüpfung von Räumen mit Genderstereotypen und Machtsystemen weniger dazu geeignet, Raumbeschreibungen in Erzähltexten aufzuspüren, zu einem späteren Zeitpunkt der Analyse kann ein solcher Ansatz aber wieder aufgegriffen werden, z. B. für eine Kategorisierung der gefundenen Raumausdrücke.

Neben dem semantischen Wertesystem, welches in der oben skizzierten feministischen Narratologie entscheidend auf Raummodelle einwirkt, gibt es in dieser Traditionslinie auch den semiotischen Ansatz, in dem der narrative Text als Zeichen- bzw. Sinnsystem[4] mit weltbildnerischer Funktion betrachtet wird (Bauer, 1997,

[4] Hier ist nicht gemeint, dass der Text an sich ein System abstrakter Zeichen (Buchstaben) ist, sondern es geht um einzelne narrative Einheiten (Begriffe, Objekte, Figuren usw.), die als Zeichen innerhalb eines Bedeutungssystems wirken.

1–3). Während die Semantisierung des narrativen Raumes bei Bal, Nünning und Warhol eher mit dem Text als Spiegel der Gesellschaft verknüpft ist, rückt Bauer die Wechselwirkung von Literatur und der Lebenswelt der Lesenden ins Zentrum der Betrachtung (Bauer, 1997, 65–68). Nach diesem Verständnis narrativen Raumes fungieren Lesende als Angelpunkt zwischen Text und Lebenswelt. Erst die Decodierung durch eine lesende Person vervollständigt den narrativen Raum (Bauer, 1997, 216) und fügt ihrer Lebenswelt eine Bedeutungsebene hinzu. Aus zwei Gründen kann dieser Ansatz bei der auf Automatisierung ausgerichteten Operationalisierung literarischer Raumdarstellung nicht weiter berücksichtigt werden. Zum einen ist das Raumverständnis nicht ausreichend klar definiert und der Begriff „Welt“ wird mal metaphorisch als Bezeichnung für das Bedeutungssystem des Textes und dann wieder konkret für die reale Welt bzw. Lebenswelt der lesenden Person (die aber unterdefiniert bleibt) verwendet. Zum anderen steht hier eine rezeptionsästhetische Deutung des Raumkonzeptes im Zentrum. Da die Konstruktion narrativer Räume durch Lesende hier nicht untersucht wird, wurde sie auch nicht in die Modellentwicklung einbezogen.

2.2.2 Topoi – Raumthemen

Die Forschung zu bestimmten literarischen Topoi ist divers, meist stark in literarhistorische Kontexte eingebunden und nur selten für ein übergeordnetes Modell generalisierbar. Literarische Topoi sind so wandelbar, dass sie zum Teil bereits innerhalb ihres literarhistorischen Kontextes unterschiedlich, manchmal sogar konträr eingesetzt werden. Zum Beispiel beschreibt Cohen (2015) den Topos der Südsee, der vor allem in der frühen Neuzeit für ein Naturparadies und die Erfahrung neuer Möglichkeiten jenseits des Strebens nach Fortschritt stand. Derselbe Topos stand aber auch für die Strapazen einer Bezwingung des Ozeans auf einer Überfahrt von der „alten“ in die „Neue Welt“ (Cohen, 2015, 371). Weitere Topoi, die literarhistorisch bedeutsam, aber nicht immer einheitlich codiert sind, sind das Mittelmeer (Janka, 2015, 301 ff), Athen als Zentrum einer mehr oder weniger demokratischen antiken Macht (Gödde, 2015, 312 ff), Rom als sich ausdehnendes Imperium (Teuber, 2015, 324 ff), der ritterliche Hof, der oftmals in machtpolitischen Krisen dargestellt wird (Jing, 2015, 335 ff), die Kathedrale als Raum der imposanten Größe und des Sakralen (Lechtermann, 2015, 344 ff) und die „Neue Welt“ (Ehrlicher, 2015, 355 ff). In all den hier aufgeführten Untersuchungen bestimmter Topoi wird deutlich, dass eine semantische oder semiotische Deutung bestimmter Räume zwar werkübergreifend möglich ist, dass eine bestimmte Deutung aber immer nur innerhalb vergleichsweise kleiner Korpora konsistent bleibt. Lechtermann (2015) stellt für den Kathedralen-

Topos sogar fest, dass jedes einzelne literarische Werk eigene Raumfigurationen inszeniere (Lechtermann, 2015, 345).

Ein singulärer Status innerhalb der literaturwissenschaftlichen Forschung zum Raum kommt Bachtins *Chronotopos* (1975) zu. Bachtins Abhandlung wurde vielfach rezipiert und das Konzept des Chronotopos häufig neben andere literaturwissenschaftliche Raumkonzepte gestellt. Dabei konnte es allerdings nicht wesentlich weiter entwickelt oder in ein Gesamtmodell narrativen Raumes integriert werden[5]. Da Bachtin in seiner Abhandlung vor allem bestimmte Chronotopoi als raumzeitliche Themen beschreibt, soll er hier im Zusammenhang mit Raumthemen betrachtet werden.

Bachtins Chronotopos-Begriff basiert auf der Idee der Raumzeit, die Einstein in seiner Relativitätstheorie entwickelte. Die Grundannahme ist, dass die Zeit eine vierte Dimension des Raumes ist und dass Raum und Zeit darum nicht losgelöst voneinander betrachtet werden können (Bachtin, 2008, 7). Für Bachtin muss eine literarische Analyse des Raumes also stets unter raumzeitlichen Gesichtspunkten durchgeführt werden. Bachtin betrachtet den Chronotopos als eine Form-Inhalt-Kategorie (Bachtin, 2008, 7). Sein primäres Interesse richtet sich nicht auf chronotopische Mikroeinheiten, wie einzelne Erwähnungen von Orten zu einem bestimmten Zeitpunkt der Erzählung oder hochfrequente Nutzung von Raumausdrücken. Stattdessen betrachtet er übergeordnete Ortskategorien in ihrer Historizität. Statt chronotopische Strukturen systematisch aufzudecken, beschränkt sich Bachtin darauf, einzelne herausragende Chronotopoi als übergeordnete literarische Raumzeit-Phänomene auszumachen und diese diachron in einer Betrachtung des europäischen Romans nachzuweisen. Ein besonderer Schwerpunkt liegt dabei auf der Betrachtung einiger Werke Rabelais'. Da Bachtin den Begriff Chronotopos für so unterschiedliche Phänomene wie „Abenteuerzeit“ (Bachtin, 2008, 12), „fremde Welt“ (Bachtin, 2008, 25), „Folklore“ (Bachtin, 2008, 30), „Begegnung“ (Bachtin, 2008, 182) und „Idylle“ (Bachtin, 2008, 160) verwendet, entsteht keine stringente, direkt übertragbare Analysekategorie.

Sowohl die anfangs genannten Untersuchungen als auch Bachtins *Chronotopos* und noch weitere raumthematische Untersuchungen – von denen es weit mehr gibt als hier erwähnt werden können und denen häufig auch die Generalisierbarkeit fehlt, die für diese Studie von zentraler Bedeutung ist – vereint die Grundannahme, dass narrativer Raum nicht unbedingt durch einzelne Worte oder Phrasen konstruiert wird. Zum Topos der Südsee gehört z. B. weit mehr als das Wort „Südsee“ und verwandte Ausdrücke wie „Strand“ oder „Meer“. Stattdessen sind es eher semantische

[5] Zoran integriert in seine Betrachtungen zwar den Begriff chronotopisch, nutzt ihn aber, anders als Bachtin, nicht für die Typologisierung bestimmter raumzeitlicher Motive.

Felder, ganze Wortgruppen, in denen räumliche und nicht-räumliche Ausdrücke zusammen wirken. Solche Wortgruppen können automatisch generiert werden und somit kann die Idee, dass Raumthemen nicht in einzelnen Wörtern, sondern Wortfeldern sichtbar werden, sehr gut in eine computergestützte Untersuchung einbezogen werden. Auch der diachrone Aspekt, der angesprochen wird, ist für die Betrachtung des hier vorliegenden Korpus' von zentraler Bedeutung und findet darum in der Modellentwicklung ebenfalls Berücksichtigung.

2.2.3 Raum als Strukturphänomen von Texten

In der post-strukturalistischen Tradition wurde der Versuch unternommen, Raum mit verschiedenen Analysekategorien zu verknüpfen und so besser aus zu differenzieren. Als einer der ersten stellte Chatman (1978) fest, dass zwischen story-space und discourse-space[6] unterschieden werden könnte. Außerdem entwickelt er fünf Parameter für die Analyse von story-space (scale, contour, position, degree, clarity) (Chatman, 1993, 97–98), die allerdings alle sehr Film-spezifisch sind, da sie den Bezug zu Kamera, Licht und anderen Einstellungen mit einbeziehen. Discourse-space umfasst nach Chatman alles, was als räumlicher Rahmen einer Erzählung gedeutet werden kann (Chatman, 1993, 102). In verbalen, nicht-filmischen Narrativen werde hier der Fokus auf das gesetzt, was Lesende auf abstrakte Weise als mentale Räume „sehen" können (ebd.). Darüber hinaus verdeutlicht Chatman auch, wie implizit Hinweise auf den Raum einer storyworld sein können. In dem Beispiel einer Figur, die eine 200 Pfund schwere Hantel mit einer Hand tragen kann, sieht er die Information, dass ihr Bizeps vergleichsweise groß sein muss (ebd). Allerdings ist diese Information auf der Textebene nur sehr vage verknüpft (über das Objekt der Hantel und die Hand als Teil eines Subjekts), sodass sie nur über diese computerlesbar wird. Ein wichtiger Unterschied zwischen filmischen und verbalen Darstellungen von Raum ist für Chatman, dass nicht-filmische Erzähltexte auch im „Nirgends" also ohne Raum erzählt werden könnten (Chatman, 1993, 106), eine

[6] Nach Todorov, der das Begriffspaar aus dem Russischen (Fabula / Sjuzet) ins Französische übersetzte, bezeichnet der Begriff „histoire" den Inhalt einer Erzählung, der sich aus Ereignissen, Figuren und ganz allgemein der Realität einer Geschichte zusammensetzt, die auf die eine oder andere Weise erzählt werden kann. Diese Erzählweise ist für Todorov der „discours" eines Narrativs. (Todorov, 1966, 132) Chatman übersetzte nun seinerseits die Begriffe ins Englische und verknüpfte sie darüber hinaus mit der Saussure'schen Matrix von Ausdruck und Inhalt auf der einen und Substanz auf der anderen Seite (Chatman, 1993, 22). Story ist bei Chatman – vereinfachend ausgedrückt – der Inhalt einer Erzählung, und Discourse der Ausdruck (Chatman, 1993, 27).

These, die hier auf dem Prüfstand steht, da auch ein „Nirgendwo“ durch die Manifestation innerhalb einer Erzählwelt zu einem „Irgendwo“ werden müsste, in dem sich die Figuren aufhalten können. Bei Chatmans Ausführungen zur Differenzierung zwischen Figur und Setting (Chatman, 1993, 139) wird deutlich, dass diese beiden Konzepte im Hinblick auf ihre Räumlichkeit nicht klar voneinander getrennt werden können. Chatman versucht einen Weg zu finden, zwischen (Haupt-)Figuren zu unterscheiden, die nicht Teil des Settings sind, sondern sich von diesem abheben, und solchen (nebensächlichen Figuren), die zum Setting gehören (Chatman, 1993, 139). Die hier durchgeführte Korpusanalyse zeigt hingegen, dass eine solche Trennung nicht dazu beiträgt, sämtliche Aspekte des literarischen Raumes im Blick zu behalten. Figuren sind unabhängig von ihrer Bedeutung für einen Erzähltext ein wichtiger Bestandteil literarischen Raumes und prägen diesen ganz erheblich (wie in Kapitel 8 ausführlich beschrieben wird). Statt zwischen Figuren mit Charakter und Figuren ohne (die darum nach Chatman eher als Teil des Mobiliars klassifiziert werden sollten (Chatman, 1993, 140) zu unterscheiden, bietet es sich eher an, zwischen der räumlichen Komponente von Figuren (z. B. ihre Körperlichkeit, ihre Wahrnehmung von Raum) und ihrer charakterlichen Ausgestaltung zu differenzieren. Auch Chatman erkennt, dass es nicht unbedingt sinnvoll ist, Charaktere nach Wichtigkeit zu klassifizieren, sondern besser, ihre Eigenschaften im Hinblick auf die Erzählung zu betrachten. Er kommt dabei allerdings nicht bis zu dem Punkt, an dem Räumlichkeit eines dieser möglichen Features ist (Chatman, 1993, 141). Interessant ist auch Zorans Ansatz, Raum im Text auf drei Ebenen anzusiedeln: einem topografischen Level, einem chronotopischen[7] und einem textuellen Level (Zoran, 1984, 315). Auf der textuellen Ebene wird narrativer Raum in einen Bedeutungszusammenhang mit dem Erzähltext eingebettet, während auf chronotopischer Ebene Raum zwar noch mit der Handlung, aber nicht mehr unbedingt mit dem Text als solchem verknüpft ist. Die topographische Ebene ist schließlich vollkommen unabhängig vom Text und gleicht einem räumlichen Modell, welches auch ohne den Text gedacht werden kann (Zoran, 1984, 315). Zoran versieht also das Phänomen des narrativen Raumes mit unterschiedlichen Abstraktionsgraden vom Text. Dieser vertikal ausgerichtete Ansatz ist nicht der einzige, den Zoran entwirft, um die Struktur narrativen Raumes klarer herauszuarbeiten. Analog zu den drei Leveln entwickelt er drei räumliche Einheiten, die für einen Text von Bedeutung sind. Er unterscheidet zwischen „total space“, „spatial complex“ und „spatial unit“ (Zoran, 1984, 322). Am Beispiel der spatial units macht Zoran deutlich, dass diese auf allen

[7] Den Begriff des Chronotopischen setzt Zoran in seiner buchstäblichen Bedeutung als „raumzeitlich“ ein, statt wie Bachtin (vgl. Abschnitt 2.2.2 damit vor allem raumzeitliche Motive zu bezeichnen.

drei Leveln des narrativen Raumes eingesetzt werden können. Die Phänomene, die in dieser computergestützten Analyse Berücksichtigung finden, sind vor allem auf Ebene der spatial units anzusiedeln, zum Teil werden auch Raumkomplexe sichtbar gemacht. Die vorliegende Analyse stützt Zorans Idee der Abstraktionsgrade, denn sie bestätigt, dass sich spatial units auf der textuellen Ebene finden und dass die Einheit des spatial complex im Text angelegt ist, gleichzeitig aber darüber hinaus geht und damit auf die chronotopische Ebene einwirkt. Wie die Topographie eines Erzähltextes sich beim Lesen ausgestaltet, also welchen Gesamtkomplex sie am Ende des individuellen Leseprozesses ausmacht, darüber kann mit der hier gewählten Methode nichts herausgefunden werden.

Obwohl sie in vielen Punkten eher an die literaturwissenschaftliche Tradition anknüpft, die Raum auf der Bedeutungsebene von Texten untersucht, hat Mieke Bal (1985) auch einige strukturelle Differenzierungen zur Debatte des literarischen Raumes beigetragen. So findet sich in Ansätzen bei ihr, was Zoran in seinen Betrachtungen bewusst ausspart, eine Verknüpfung von Raum und narrativer Funktion (Zoran, 1984, 335). Sie unterscheidet zwischen Raum als Teil des Rahmens einer Erzählung, Raum als beschreibendem Element und Raum als emotionalem Bezugspunkt für Figuren (Bal and van Boheemen-Saaf, 1985, 136–137). Darüber hinaus benennt sie – ähnlich wie Genette (Genette und Knop, 1998, 21 ff) es für die Zeit und Chatman für story-space (s.o.) getan hat – fünf Parameter, die narrativen Raum bestimmen: Determination, Repetition, Akkumulation, Transformation und Relation von Räumen zueinander (Bal and van Boheemen-Saaf, 1985, 139). Bal fügt dieser Unterscheidung noch die Differenzierung zwischen „place“ und „space“ hinzu. Den Begriff place definiert sie als Überbegriff für topologische Positionen, an denen sich Charaktere befinden und Events stattfinden können (Bal and van Boheemen-Saaf, 1985, 133). Space entstehe, wenn solche Orte mit der sinnlichen Wahrnehmung (insbesondere über die drei Sinne des Sehens, Hörens und Fühlens) der Charaktere verknüpft würden (Bal and van Boheemen-Saaf, 1985, 133).

Auch in der rhetorisch interessierten Erzählforschung werden die Funktionalitäten narrativen Raums hervorgehoben. Setting wird als zentrale räumliche Kategorie ausgemacht, die z. B. geographische Begebenheiten und die darin befindlichen Objekte beinhalte, und damit so umfassend sei, dass es nahezu automatisch zum Hintergrund eines Erzähltextes werde (Phelan, 2012, 85). Phelan und Rabinowitz machen drei Hauptfunktionen des Settings aus: die synthetische oder formale, die mimetische und die thematische (Phelan, 2012, 85). Auf formaler Ebene gibt das Setting einem Erzähltext einen Rahmen und legt die Bedingungen des Erzählverlaufs fest (Phelan, 2012, 86). Die mimetische Funktion des Settings ist im Grunde eine beschreibende, die die Aufgabe hat, einer lesenden Person als Fenster zur realen Welt zu dienen (Phelan, 2012, 87). Die thematische Funktion des Settings

umfasst schließlich alle Hinweise auf symbolische oder semantische Aufladung von Räumen. Die hier aufgezeigten analytischen Kategorien zielen allerdings weniger darauf ab, interne Strukturen von Erzählungen besser fassen zu können, wie es die Ansätze von Chatman, Zoran und Bal tun. Stattdessen können mit Hilfe dieser Funktionen des Settings die Wirkungszusammenhänge untersucht werden, seien diese textimmanent wie bei der synthetischen Funktion oder rezeptionsästhetisch wie bei der mimetischen und der thematischen Funktion. Darum werden diese Funktionen in der folgenden Analyse nur zum Teil sichtbar. Die synthetische oder formale Funktion wird zwar voll berücksichtigt, die mimetische und thematische Ebene aber nur in den Grenzen dessen betrachtet, was im Text explizit oder implizit angelegt ist. Die hier dokumentierte Studie zeigt allerdings, dass gerade die thematische Funktion von Raumdarstellungen sich durchaus auch textimmanent nieder schlägt und eine wichtige, wenn auch weniger häufige als die synthetische, Dimension von Raumdarstellungen im Text ausmacht.

Bei allen oben genannten Ansätzen geht es nicht primär darum, Raumreferenzen in Erzähltexten ausfindig zu machen, sondern bereits herausgearbeitete Fundstücke zu klassifizieren und zu kategorisieren. Einige der konzeptionellen Ansätze wie z. B. die Unterscheidung zwischen place und space nach Bal und zwischen konkreten, im Text befindlichen Raumeinheiten und abstrakteren Raummustern wie sie sich sowohl bei Zoran als auch bei Phelan und Rabinowitz finden, wurden in die theoriegeleitete Modellbildung literarischen Raumes einbezogen. Andere, eher konkrete Fundstücke in Erzähltexten klassifizierende Kategorien wie Chatmans fünf Raumparameter oder Bals Unterscheidung zwischen Determination, Akkumulation, Transformation und Relation werden zum Teil bei der Auswertung der Korpusanalysen wieder aufgegriffen. Bei der hier dokumentierten Analyse hat sich gezeigt, dass gerade die Betrachtungen der Strukturen, die durch narrative Raumdarstellungen entstehen, von der computergestützten Analyse profitieren kann.

2.2.4 Literarischer Raum als Raumkonstruktion zwischen Text und Leser*in

Der grundlegende Gedanke, dass literarischer Raum sich im Zusammenspiel von Text und Leser*in entwickelt, ist zentral für die rezeptionsästhetische literaturwissenschaftliche Forschung (so z. B. in Butor, 1968, passim), für die kognitivistisch interessierte Narratologie (vertreten u.a. von Herman, 2012, passim) und auch für die Possible-Worlds-Theorie (z. B. Pavel, 1986, passim, Ryan, 1980, passim, Ronen, 1994, passim, Doležel, 2009, passim). Grundlegend ist für diese Ansätze die Idee, dass Leser*innen sich auf eine mentale Reise zum Ort des Romans begeben (Butor,

1968, 79; Herman, 2012, 98). Auf mentalen Karten werden die im Erzähltext enthaltenen und sprachlich aufbereiteten Informationen festgehalten (Herman, 2012, 99). Ähnlich wie bei Bauer in seinem semiotischen Ansatz (vgl. Abschnitt 2.2.1) wird auch hier angenommen, dass mittels Sprache beim Lesen ein narrativer Raum erschaffen wird (Herman, 2012, 99–100). Im Zentrum der Possible-Worlds-Theorie steht aber auch die Frage nach dem ontologischen und psychologischen Status von Erzählungen und dem Unterschied von Fakt und Fiktion (Pavel, 1986, 75 ff). Da sowohl die Frage nach der (Re-)Konstruktion narrativen Raumes durch Lesende als auch die nach dem ontologischen Status narrativer Räume über eine textimmanente Betrachtung des Raumes in der Literatur hinaus geht, sind in diesen Forschungszweigen leider keine neuen Aspekte für die Modellierung eines operationalisierbaren Raummodells zu finden. Sie wurden darum nicht in diese Studie einbezogen.

2.2.5 literaturwissenschaftliche Aspekte des Raummodells

Aufgrund der vorausgehenden zusammenfassenden Betrachtungen der Forschung zum literarischen Raum konnten bereits zentrale Aspekte für vier der Kategorien des zu Beginn dieses Kapitels beschriebenen und im fuzzy-set-Modell narrativen Raumes integrierten Kategoriensystems ausgearbeitet werden. Konkrete Raumreferenzen wie sie z. B. Piatti und Dennerlein beschreiben und die von Zoran als spatial units benannt werden, sind zentraler Bestandteil der Kategorie „Orte“. Räumliche Relationen manifestieren sich zwar auch konkret im Text, wirken aber definitionsgemäß nicht allein, sondern nur im Zusammenhang mit anderen, nicht primär räumlichen Faktoren des Textes. Relationen können z. B. zwischen Figuren oder Objekten bestehen, aber sie bestehen nicht per se. Trotzdem handelt es sich auch hier um spatial units auf der Textebene. Auch hierfür hat Dennerlein bereits konkrete Referenzen vorgeschlagen. Raumthemen und Chronotopoi können wie bereits beschrieben als Wortfelder konkretisiert werden, die mit Raumausdrücken in Verbindung gebracht werden können, sodass die Forschung zu Topoi, wie sie in Abschnitt 2.2.2 zusammengefasst wurde, auf dieser Ebene eingearbeitet worden ist. Nach Zoran erreichen wir mit den Raumthemen die chronotopische Ebene und betrachten nicht mehr nur spatial units, sondern Raumkomplexe. Am wenigsten explizit sind Hinweise auf den Raum, wie z. B. ein Verweis auf die Gestalt einer Figur, der zwar impliziert, dass diese Raum einnimmt, die Raumdarstellung aber nicht in den Fokus rückt. Bei einem Ausdruck wie „sie war von zierlicher Gestalt“ gibt es z. B. einen vagen Hinweis auf den Raum (die Figur nimmt vergleichsweise wenig Raum ein), dieser ist aber kaum mehr als eine Assoziation von vielen. Ästhetische Aspekte, gesellschaftliche Normen und auch Implikationen von Kraft und evtl. sogar (mangelnder)

Durchsetzungsfähigkeit schwingen hier mit. Es entsteht ein komplexer Assoziationszusammenhang, der auch Raum enthält, der aber relativ vage im Text verankert ist. Diese Ebene ist zwar nach Zoran immer noch chronotopisch, allerdings abstrakter und impliziter räumlich als Raumthemen es sind. Obwohl die Verbindung von Wortebene und räumlicher Dimension hier vage ist, kann, solange diese Verbindung besteht, auch die computergestützte Methodik fruchtbar eingesetzt werden. In den abstrakteren Raumkategorien der Raumthemen und -hinweise wurde auch die raumsemantische Traditionslinie von Bachelard über Lotman bis hin zu Ansätzen aus der feministischen Narratologie bei der Modellierung einbezogen. Konkret benannte Raumsymbole wie z. B. Fenster und Treppen (vgl. Abschnitt 2.2.1) werden hier zu Indikatoren der größeren Raumeinheit der Raumthemen.

Metaphern, die Raumausdrücke nutzen, um eine nicht-räumliche Bedeutung zu transportieren, wurden nicht berücksichtigt. Grund hierfür ist, dass Raummetaphern zwar Raumausdrücke enthalten, sie beschreiben aber nicht räumliche Aspekte, sondern implizieren nicht-räumliche Bedeutungen. Das heißt, dass später, bei der Implementierung des Modells, ein Arbeitsschritt zur Extraktion der metaphorisch verwendeten Raumausdrücke eingefügt werden muss. Der Metaphern-Begriff wird hier nicht als Oberbegriff für semantisch-kulturelle Aufladung verwendet, sondern als Bezeichnung für Phrasen, die Raumausdrücke enthalten aber keine Raumbeschreibungen sind, also für eine uneigentliche Wortverwendung. Wie bereits erwähnt, wurden semantisch-kulturelle Implikationen von Raumausdrücken dagegen sehr wohl ins fuzzy-set-Raummodell einbezogen, solange sie eine zumindest implizite Verknüpfung mit räumlichen Aspekten der storyworld haben.

Anschließend an die literaturwissenschaftlichen Betrachtungen wird dieser Studie ein relativ weiter Raumbegriff zu Grunde gelegt. Ausgehend vom phänomenologischen Raum werden Orte und Räume hier als wahrnehmbare Größen verstanden, die sowohl subjektiv erfahrbar als auch objektiv kartographierbar sein können. Betrachtet wird hier die Darstellung von Raum, d. h. Ausdrücke und Beschreibungen, die Orte und Räume implizit oder explizit aufrufen. Nachdem festgelegt wurde, dass ein phänomenologisches Raumverständnis für diese Arbeit grundlegend ist, und nachdem die literaturwissenschaftlichen Betrachtungen bereits gezeigt haben, wie die Darstellung von Raum umgesetzt sein kann, schließt sich die Frage an, welche Phänomene genau zu den räumlichen gezählt werden können. Diese Frage kann nicht auf Basis literaturwissenschaftlicher Forschung allein beantwortet werden, weshalb ich hier einen wissenschaftsgeschichtlichen Überblick zur Raumforschung außerhalb der Literaturwissenschaften anschließe.

2.3 phänomenologische Grundlagen der Raumdarstellung für das fuzzy-set-Modell

Um das Modell narrativen Raumes weiter zu konkretisieren, wurden ergänzend Ansätze aus anderen Disziplinen ausgewertet. Im nun folgenden Abschnitt liegt der Fokus auf Betrachtungen, die einen phänomenologischen Raumbegriff stützen. Denn wie oben bereits angedeutet, wird der konkrete, wahrnehmbare Raum hier als Grundlage für den erzählerisch dargestellten Raum verstanden. Es geht also um ein erfahrungsbasiertes mimetisches Raumkonzept, das im theoretischen Modell implementiert wird. Grundlage ist zwar die Raumerfahrung (sowohl auf Seite der Rezipierenden als auch auf Seite der Figuren in der fiktionalen Welt), diese wird aber im Erzähltext stets mittelbar dargestellt, sei es als Erfahrung der Figuren oder als bloße Vorstellung derselben – der Raum im Erzähltext ist nie physisch. Aus diesem Grund wird hier auch nicht zwischen tatsächlichen und mentalen bzw. inneren Räumen der Figuren unterschieden. Anders als Schmid (Schmid, 2008, 24), der für seinen Eventbegriff klar zwischen vorgestellten und innerhalb der storyworld tatsächlich stattfindenden Zustandsveränderungen unterscheidet und nur Letztere als events definiert, beziehe ich alle erwähnten Räume in meine Betrachtungen ein, seien sie nun vorgestellt oder tatsächlich Teil der storyworld, seien sie innerhalb oder außerhalb einer Figur verortet (mehr dazu in Kapitel 3). Am Ende dieser Betrachtungen wird die Trennschärfe dessen, was hier als räumliche Dimension von Erzählungen betrachtet wird, erhöht. Im Folgenden werden drei Traditionslinien der Raumbetrachtung zusammengefasst, die sich vor allem durch den Fokus auf unterschiedliche Aspekte des Raumes voneinander unterscheiden. Damit geht auch einher, dass Raum unterschiedlich abstrakt betrachtet wird. In Abschnitt 2.3.1 steht die Entwicklung eines Objekt-zentrierten und damit sehr konkreten Raumbegriffes im Zentrum. Darauf aufbauend wird in Abschnitt 2.3.2 ein relationales Raumverständnis beschrieben, das immer mindestens zwei konkrete Größen und eine abstrakte Verbindung dazwischen benötigt. Schließlich wende ich mich in Abschnitt 2.3.3 noch der Betrachtung sozialer und kultureller Räume zu, die sowohl konkrete Objekte als auch Relationen dazwischen und darüber hinaus noch abstrakte Wertmodelle einbeziehen.

Wissenschaftliche Traditionslinien der Betrachtungen zum Thema Raum sind zwar divers und haben viele unterschiedliche Ansatzpunkte. Die hier betrachteten Schlüsselpositionen lassen sich aber grob in drei Richtungen unterscheiden. Diese sind hauptsächlich inhaltlich begründet, zeigen aber auch eine zeitliche Entwicklung. Während frühe Ansätze häufig Objekte als räumliche Elemente in den Fokus stellten (absoluter Raumbegriff), wurde Raum später zunehmend relational, also entweder in einer Subjekt-Objekt-Relation oder als System aus Beziehungen zwischen Objekten (relationaler Raumbegriff), betrachtet. Mit dem spatial turn in der

zweiten Hälfte des 20. Jahrhunderts, der vor allem die Sozial- und Kulturwissenschaften betraf (Günzel, 2009a, 10–11 und Dennerlein, 2009, 6), gab es zunehmend Ansätze, die Objekte und deren Relationen zueinander in ihrer gesellschaftlichen und kulturellen Bedeutung betrachteten (Günzel, 2009b). Unberücksichtigt bleiben hier Ansätze, die von der phänomenologischen Betrachtung komplett absehen und Raumbegriffe lediglich in ihrer metaphorischen Bedeutung nutzen und untersuchen oder die Fragestellungen nach dem ontologischen Status von Raum nachgehen, wie z. B. der Fall bei (Hegel, 1979 und Heidegger, 2006).

2.3.1 Objekt-zentrierter Raumbegriff

Erste Ansätze zu Objekt-zentrierten Raummodellen wurden bereits in der Antike entwickelt. Bei Aristoteles finden sich sowohl in der *Physik* als auch in der *Metaphysik* Auseinandersetzungen mit unterschiedlichen Ideen zu den Fragen, ob es Raum gibt und wenn ja, was Raum ist[8]. Er gelangt zu dem Schluss, dass es einen Raum geben muss, da etwas seine Lage, also seinen Ort, wechseln kann (Aristoteles, 1995, 82). Wenn dort, wo einmal Luft war, nun Wasser sein kann, dann muss der Ort etwas sein, das weder direkt zum einen noch zum anderen gehöre (Aristoteles, 1995, 82). Die sechs Richtungen (oben, unten, rechts, links, vorne, hinten) seien für den Raum nicht relativ wie für den Menschen, für den sie sich je nach Lage veränderten, sondern festgelegt (Aristoteles, 1995, 82). Die Richtungen des Raumes ergeben sich aus seinen physikalischen Gesetzmäßigkeiten, wie z. B., dass Leichtes nach oben strebt (Aristoteles, 1995, 83). Mit seinem Schluss, dass der Raum so ähnlich wie ein Gefäß sei (Aristoteles, 1995, 86), legt Aristoteles den Grundstein für die später bedeutsam werdende Idee des Container-Raumes. Während der Körper sich beim Verständnis vom Raum als Container[9] allerdings innerhalb dieses Raumes bewegt, ist die Gefäß-Idee des Aristoteles sehr eng an den Körper gebunden; Raum sei dessen direkte, unmittelbare, unbewegliche Grenze (Aristoteles, 1995, 93).

Obwohl die Idee des Raumes als unmittelbarer Grenze des Körpers aus heutiger Sicht überholt wirken mag, ist die Grundidee der engen Verknüpfung von Raum und Körper bzw. Objekt es nicht. Neben Beispielen wie Bachelard, Lotman und Tuan, die bereits in Abschnitt 2.2 genannt wurden, geht z. B. auch Flusser (2006) davon

[8] Z. B. grenzt Aristoteles sich von Hesiod ab, der für einen apriorischen Raumbegriff argumentiert hat (Aristoteles, 1995, 83) und verfolgt dafür Ideen von Zenon weiter (Aristoteles, 1995, 88).

[9] Dass die Idee des Raums als Container bis heute fest mit einem Alltagsverständnis des Phänomens verknüpft ist, stellen sowohl Löw und Sturm in ihrer *Raumsoziologie* (Löw und Sturm, 2016, 2) als auch Dennerlein in ihrer *Raumnarratologie* (Dennerlein, 2009, 60) fest.

aus, dass die körperliche Erfahrung des Menschen zentral für sein Raumempfinden ist. Die Benennung eines jeden Gegenstandes und jeder Figur einer Erzählung kann als Hinweis auf eine räumliche Dimension gelesen werden. In Bezug auf die Erzähltextanalyse ist dies teilweise problematisch. Denn wenn jede Erwähnung einer Figur als Hinweis auf den Raum interpretiert wird, so ergibt sich eine starke Überlagerung mit dem Konzept der Figur. Ähnliches gilt für die Bewegung, die für das aristotelische Raumverständnis ebenfalls zentral ist. Wenn jede Beschreibung von Bewegung in einem Erzähltext als Hinweis auf den Raum angenommen wird, so fehlt die Trennschärfe zur Kategorie der Zeit. Aus diesem Grund werden zwar Figurennennungen und sprachliche Markierungen von Bewegung nicht zwingend als Hinweise auf den Raum betrachtet, wenn sie allerdings als räumliche Variablen gekennzeichnet werden (wie z. B. indem sie einen Wechsel von Positionen bezeichnen), so wird dies durchaus in dieser Studie berücksichtigt. Auch der Hinweis auf die potentielle Beweglichkeit aller Objekte und die damit verbundene räumliche Flexibilität wird konzeptionell in die Modellierung von Raum einbezogen. Neben der Beweglichkeit von Objekten und dem damit verbundenen potentiellen Positionswechsel sind die Richtungen rechts-links, oben-unten, vorne-hinten als subjektive und objektive Relationen von Raum von Bedeutung. Die von Aristoteles erwähnten (s.o.) absoluten Richtungen deute ich als Himmelsrichtungen und beziehe sie als Positionsbestimmungen (Kategorie: „Orte“) ins Raummodell ein. Positionen und Richtungen können durch bestimmte Worte und Phrasen ausgedrückt werden und sind somit operationalisierbar. Positionen und Lageveränderungen können auf Ebene der konkreten Raumreferenzen ausgedrückt werden, ebenso wie die objektiven Richtungen, wie z. B. oben und unten. Subjektive Richtungen können relational ausgedrückt werde, wie z. B. in den Phrasen „links von ihr“ oder „direkt vor ihnen“[10].

Während Aristoteles beim Objekt als kleinster Einheit seines Raumverständnisses ansetzt, rückt die euklidische Geometrie den Punkt als unzerteilbares, kleinstes räumliches Element ins Zentrum. Das auf dieser Idee aufbauende Raumverständnis wird bis heute als „euklidischer Raum“ bezeichnet, obwohl nicht ganz klar ist, ob Euklid nur zahlreiche Ansätze der Platoniker zusammenführte vgl. Reidemeister (1940, 6) und Becker und Patzig (1966, 18–20) und daraus ein umfassendes und bis

[10] Dennerlein hat die Raumrichtungen „Osten“ und „Westen“ als nicht näher spezifizierte „weitere Konkreta“, die objektiven Raumrichtungen „oben“ und „unten“ als dimensionale Positionierung, die deiktisch ausgedrückt wird, und die subjektiven Richtungen als deiktische Ortsangaben in ihr Modell einbezogen und somit für die Konkretisierung von Raum in Narrativen fruchtbar gemacht (Dennerlein, 2009, 208–209), darum wird dieser sehr feingliedrige Ansatz bei der späteren Operationalisierung der angelegten Ebenen des Raummodells wieder herangezogen.

heute zumindest für die Schulgeometrie grundlegendes vgl. Waerden (1956, 321) und Filler (1993, I) Werk der Geometrie machte oder ob er tatsächlich auch eigene Ansätze in seinen *Elementen* festhielt (Becker und Patzig, 1966, 18–20). In den *Elementen* Euklids finden sich zahlreiche aufeinander aufbauende Begriffsdefinitionen, Axiome und Postulate (Simons, 2015, 272), die für einen Objekt-zentrierten, absoluten Raumbegriff bedeutsam sind. Basis eines jeden Objekts ist der Punkt, aus vielen Punkten können Linien, aus vielen Linien Flächen (Euclides, 1971, 1, Buch I Definitionen 1–5) werden, die wiederum Körper begrenzen können, die ihrerseits durch die Ausdehnung in Länge, Breite und Tiefe bestimmt sind (Euclides, 1971, 315, Buch XI Definitionen 1 und 2). Für unterschiedliche Arten von Flächen und Körpern werden außerdem mathematische Gesetzmäßigkeiten festgehalten, wie z. B. dass die Winkelsumme eines Dreiecks immer 180° beträgt (Euclides, 1971, 23, Buch I §32). Entscheidend für den euklidischen Raum ist, dass es sich um einen idealen Raum handelt (Reidemeister, 1940, 12–13). Es gilt das Gesetz der Widerspruchsfreiheit (Reidemeister, 1940, 12–13). Obwohl viele der euklidischen Axiome und Postulate mathematisch bewiesen werden können, greift im Zweifelsfalle die definitorische Setzung in einer Art Zirkelschluss: Da Parallelen als zwei Linien definiert sind, die sich nie treffen (Euclides, 1971, 2, Buch I Definition 23), können sich Parallelen nicht treffen. Die euklidische Geometrie ist also zutiefst theoretisch. Darum gilt sie heute auch als überholt bzw. als durch den Mathematiker Carl Friedrich Gauß überwunden und durch die Idee eines gekrümmten Raumes ersetzt (vgl. Filler, 1993, 163–164 und 213–216). Für diese Studie ist es vor allem die grundsätzliche Idee, dass Raum aus einzelnen Einheiten zusammengesetzt ist, die so zusammengefügt werden können, dass sich nach und nach mehrere Dimensionen ergeben, die fruchtbar gemacht werden kann. Als Basis einer literarischen Raumdarstellung können Orte und örtliche Objekte[11] (wie z. B. das Haus) als kleinste Einheiten definiert werden. Aus mindestens zwei Orten und / oder Objekten ergibt sich eine Relation

[11] Orte, Relationen und Topographien in Erzähltexten funktionieren zwar ähnlich wie Euklids *Elemente* in der Hinsicht, dass sie zusammengefügt zu größeren Einheiten werden können, sie haben aber grundverschiedene Eigenheiten. So sind Orte im Gegensatz zu Punkten nicht unteilbar, sondern haben bereits eine Ausdehnung, die reduziert oder erweitert werden kann. Relationen sind, anders als Linien, weder zweidimensional noch in irgendeiner Weise physisch. Sie verbinden zwar Orte, Objekte und Figuren miteinander, aber nicht auf eine bestimmte Weise (sie können objektive Qualitäten annehmen wie z. B. Distanz, aber auch gänzlich andere, wie z. B. emotionale Verbundenheit, qualitative Mehrfachcodierung ist ebenfalls möglich, z. B. Distanz auf räumlicher und emotionaler Ebene). Auch die Topographie literarischen Raumes ist nicht flach, sondern mehrdimensional. So wird der literarische Raum letztendlich auch nicht zu einem euklidischen Raum, sondern teilt nur die Funktion der Zusammensetzbarkeit seiner Elemente und die Tatsache, dass es sich um einen theoretischen, d. h. nicht physisch manifestierten Raum handelt.

(analog zur Linie), aus mehreren Orten und / oder Objekten und / oder Hinweisen auf diese eine Art Topographie (analog zur Fläche). Eine weitere Dimension bildet die kulturelle Aufladung von Räumen. Damit ist die Basis für das mehrdimensionale Raummodell geschaffen, das zu Beginn dieses Kapitels vorgestellt wurde.

Eine wichtige Entdeckung innerhalb der Raumforschung war, dass das Universum nicht geo- sondern heliozentrisch angeordnet ist (Kopernikus, 2007, 16–21). Die Bedeutung der damit verbundenen Kopernikanischen Wende wurde als einschneidendes kulturelles Erlebnis innerhalb der Menschheitsgeschichte auch schon aus unterschiedlichen Perspektiven, wie z. B. der psychoanalytischen durch Freud oder der kulturtheoretischen durch Blumenberg analysiert (Schmidt, 2009, 297). Indem Kopernikus die Erde und damit auch die Menschheit aus dem Zentrum des Universums gerückt hat (Schmidt, 2009, 297), macht er sie zu einem beobachtbaren räumlichen Objekt unter anderen. Der Mensch in seiner Rolle als Beobachter rückt von einem zentralen an einen peripheren Ort des Universums. Durch das Infragestellen der Zentralität des Subjekts nimmt die Bedeutung von Subjekt-Objekt-Relationen zu. Gleiches gilt für die Beziehungen räumlicher Objekte zueinander, die im Fokus von Kopernikus‘ Abhandlung stehen. Damit weist Kopernikus’ Ansatz recht weit über seine Zeit hinaus und legt den Grundstein für Denkmodelle, die eigentlich erst ab dem 18. Jahrhundert ins Zentrum der Beschäftigung mit dem Raum rücken und die hier als relationale Raummodelle benannt werden. Die Kopernikanische Wende verweist aber noch auf einen weiteren Aspekt, der für die vorliegende Arbeit wichtig ist: Es ist möglich, Raum subjektiv wahrzunehmen und daraus ein Weltbild abzuleiten, das eine Kultur über Jahrhunderte hinweg prägt und trägt. So hat es das geozentrische Weltbild getan und auch die Vorstellung der Welt als Scheibe war menschheitsgeschichtlich konsistent und tragbar. Das heißt, dass ein Raumkonzept oder Raumverständnis (und damit auch ein narrativ entworfener Raum) nicht empirisch fundiert oder gar verifiziert sein muss, um als schlüssig und tragfähig angenommen zu werden. Die Kopernikanische Wende hat auch gezeigt, dass sowohl ein individuelles (durch Kopernikus selbst) als auch ein kollektives (kulturelles) Umdenken möglich ist. Diese Dualität in der Perspektive auf Raum ist immer mitzudenken und ist darum auch für ein erzähltheoretisch anwendbares Raumkonzept unverzichtbar. Die Kopernikanische Wende stützt also die These, dass narrativ entworfener Raum in keiner Weise einem geographischen Raum entsprechen oder auch nur bis ins letzte Detail konsistent sein muss, ohne dass dies ein Problem für das Verständnis der storyworld insgesamt darstellt.

Bei Aristoteles, in der euklidischen Geometrie und letztendlich auch bei Kopernikus ist die Definition räumlicher Elemente als abgeschlossener Einheiten zentral. Dagegen steht die Idee eines räumlichen Kontinuums, die von Descartes entwickelt wurde (Einstein, 2006, 95). Allerdings geht auch Descartes in seinem Raummodell,

ähnlich wie Aristoteles, sehr stark von Körpern und damit von Objekten aus. Er definiert aber nicht die Grenze dieser Körper als Raum, sondern deren Ausdehnung, sodass sie zur reinen res extensa werden (Descartes, 2007, 93 und 95). Von dieser Idee des Raumes als Ausdehnung in Tiefe, Breite und Höhe, die in ihrem Kern die euklidische Definition von Körpern aufgreift, grenzt Descartes den Begriff des Ortes ab. Im Gegensatz zum Raumbegriff, der Größe und Gestalt einer Sache bezeichne, würden als Ort eher Positionen und konkrete Lagebestimmungen benannt (Descartes, 2007, 107)[12]. Für Descartes gibt es keinen leeren Raum (Descartes, 2007, 131). Es ist aber möglich, Räume nach Qualitäten wie z. B. Festigkeit zu unterscheiden. Die Idee eines räumlichen Containers sei im Prinzip nichts anderes als die Vorstellung, dass eine räumliche Einheit von besonders standhaften Körpern, also von anders gearteten räumlichen Einheiten, umgeben sei (Descartes, 2007, 103). Da es für Descartes keinen leeren Raum gibt, bedeutet Bewegung für ihn, dass sich eine räumliche Einheit aus der unmittelbaren Nachbarschaft der einen Körper in die unmittelbare Nachbarschaft anderer Körper begibt (Descartes, 2007, 121). Dieses Prinzip ist nicht nur einleuchtender als die aristotelische Vorstellung, dass ein Körper in der Bewegung permanent seine Grenze ändert, sondern hilft auch dabei den Raum als Fluidum zu begreifen, das permanentem Wandel unterliegt. Gleichzeitig unterstützt die strikte Trennung der Begriffe Ort und Raum die Idee eines dreidimensionalen euklidischen Raummodells. Einerseits wird hier klar, dass weder einzelne Positionen noch zweidimensionale verbundene Positionen als Raum verstanden werden können, sondern nur etwas, das mindestens drei Dimensionen besitzt. Auf der anderen Seite werden Positionen hier analog zu den Punkten des euklidischen Raumes zu kleinsten Einheiten, die mit Hilfe von Relationen zweidimensional erweitert werden können. Erst die kulturelle Aufladung, die hier als Einbindung der grundlegenden Ortseinheiten (Positionen und Relationen) in die Thematik einer Erzählung gedeutet wird, vervollständigt den literarischen Raum bzw. macht ihn erst zum Raum. Drei Aspekte werden also von Descartes ins operationalisierbare Raummodell übernommen: Die Ununterscheidbarkeit zwischen Raum und nicht-Raum und die weniger binäre Differenzierung nach Raumqualitäten, die damit einhergeht, die Bewegung als Veränderung der Beziehungen eines Raum-Elements zu anderen Raum-Elementen und die strikte Trennung von Orts- und Raumbegriff, bei dem letzterer immer etwas mindestens Dreidimensionales bezeichnen muss.

---

[12] Bis heute wird die Idee des Raums als mehrdimensionales Bezugssystem, in dem Positionen und Relationen verortet werden können, wie z. B. auf einer Karte in Rückbezug auf Decartes als „Cartesianischer Raum“ bezeichnet.

Der Begriff der Gestalt, den Descartes als für räumliche Einheiten bezeichnend darstellt, beinhaltet darüber hinaus eine neue Ebene der Hinweise auf Raum als Eigenschaft von Objekten. Demnach könnten auch Adjektive wie etwa flach, spitz oder gekrümmt auf räumliche Aspekte, nämlich die Räumlichkeit von Objekten und auch von Subjekten hinweisen. Hier wird wieder die Problematik offenbar, dass das Raumkonzept sich mit Aspekten der Figurenbeschreibung überlappt. Wenn ein Hinweis auf Objekte und Subjekte, der auch implizit sein kann, ausreicht, um Raum zu beschreiben, so trägt dies nicht zur Trennschärfe eines literarischen Raumkonzepts bei. Dieses würde im Gegenteil beliebig werden. Darum musste hier beim Machine-Learning-Training, das auf die Modellierung der Raumkategorien folgte, ganz besonders darauf geachtet werden, wann in adjektivischen Beschreibungen tatsächlich Aspekte von Raum aufgerufen wurden und wann nicht. Bei Raumhinweisen, die sich auf Figuren bezogen, war ebenfalls besondere Sorgfalt nötig. Nur wo tatsächlich auf die Körperlichkeit und die räumliche Ausdehnung von Figuren referenziert wurde, wurden diese auch als Raumhinweise bewertet. Da, wie bereits erwähnt, bei Beschreibungen der Körper von Figuren Raum eher eine untergeordnete Rolle im Assoziationskomplex einnimmt, wurde diese Kategorie als weniger räumlich bewertet als die konkreteren.

2.3.2 Relationaler Raumbegriff und Anschauungsraum

Ab dem 18. Jahrhundert beginnt die Entwicklung weg vom objektzentrierten Raumbegriff hin zu einer eher an Relationen interessierten Betrachtungsweise. In seiner *Monadologie* noch vom Objekt ausgehend, entwickelt sich Leibniz zu einem der wichtigsten Vertreter des relationalen Raumes. In der *Monadologie* vertritt Leibniz ein Raumverständnis, welches dem von Descartes ähnelt. Leibniz geht ebenfalls vom ausgedehnten Objekt aus und betont die Nachbarschaft von Körpern zueinander (Leibniz und Schneider, 2002, 137). Später wird die Ausdehnung von Objekten für sein Raummodell zunehmend peripher und die Relationen rücken ins Zentrum. Dies wird besonders deutlich in seinem Briefwechsel mit Clarke, der in die Wissenschaftsgeschichte als eine der wichtigsten Auseinandersetzungen der Raumforschung eingegangen ist. Leibniz vertritt darin einen relativen Raumbegriff, während Clarke als Schüler und Anhänger Newtons für einen absoluten Raumbegriff argumentiert. In seinen insgesamt fünf Briefen führt Leibniz aus, dass der Raum relativ, ebenso wie die Zeit ein theoretisches Ordnungssystem (Leibniz, 1717d, Absatz 4) und unabhängig von Körpern, die zueinander in Relation stehen, nicht vorhanden sei (Leibniz, 1717d, Absatz 5). Für Clarke und andere Anhänger Newtons ist der Raum

hingegen absolut, unveränderlich und unabhängig von menschlicher Erkenntnis und Objekten (Clarke, 1717b, Absatz 4).

Ganz im Sinne seiner *Monadologie* stellt Leibniz in seinen Briefen an Clarke ebenfalls fest, dass es nach seinem Raumverständnis kein Vakuum gebe (Leibniz, 1717c, Absatz 2 und 1717b, Post Scriptum). Anders als z. B. Descartes beschreibt Leibniz hier einen Unterschied zwischen Raum und körperlicher Ausdehnung, was er damit erklärt, dass ein Körper sehr wohl den Raum verändern könnte, nicht aber seine eigene Ausdehnung (Leibniz, 1717a, Absatz 37). Nach der Newton'schen Auffassung sind körperliche Ausdehnung und Raum strikt voneinander zu trennen, der ausgedehnte Körper existiere lediglich im Raum (Clarke, 1717a, Absatz 38). Raum ist für Newton und seine Anhänger eine nicht veränderbare Größe (Schmidt, 2009, 298). Auf diese Größe bezieht er die Beobachtungen seiner Experimente (Schmidt, 2009, 298). Rotiert man z. B. einen halb mit Wasser gefüllten Eimer schnell um eine Achse, so sorgen Trägheits- oder Zentrifugalkräfte nicht nur dafür, dass das Wasser im Eimer bleibt, sondern auch dafür, dass dessen Oberfläche sich wölbt (Schmidt, 2009, 298). Diese Kräfte werden von Newton durch die Bewegung des Eimers gegen den absoluten Raum begründet (Schmidt, 2009, 298). Er stellt damit fest, dass vom absoluten Raum Kräfte ausgehen (Schmidt, 2009, 299). Das Beispiel der Schlussfolgerung aus diesem Eimerexperiment zeigt, wie starr das Newton'sche Raumkonzept ist. Newtons Ideen vom absoluten Raum stießen aber nicht nur bei Leibniz auf Kritik. Auch Ernst Mach zweifelte diese an (Schmidt, 2009, 299). Trägheitskräfte könnten ebenso gut zwischen unterschiedlichen Materiemassen wirken wie zwischen Materie und absolutem Raum (Schmidt, 2009, 299). Für den Menschen sei es unentscheidbar, ob das eine oder das andere wahr sei (Schmidt, 2009, 299). Die Trägheit als Wechselwirkung von Massen wurde von Einstein als „Mach'sches Prinzip" bezeichnet (Schmidt, 2009, 300), auf das er in der Entwicklung der Relativitätstheorie vage Bezug nahm (Schmidt, 2009, 300).

Leibniz kommt zu einem dynamischeren Raummodell als Newton, für das die Idee zentral ist, dass ein Subjekt Raum über die Relation zu anderen Objekten, d. h. Objekten an anderen Orten wahrnimmt (Leibniz, 1717a, Absatz 47). Verändern sich die Positionen, so verändern sich auch die Relationen und damit der Raum. Seine Definition von Ort ist also die von einem Subjekt oder Objekt eingenommene Position, die in Differenz und Relation zu anderen Positionen bzw. zu anderen Objekten, die diese einnehmen, bestimmt wird (Leibniz, 1717a, Absatz 47). Raum definiert er als mehrere Orte zusammengenommen (Leibniz, 1717a, Absatz 47), d. h. Raum ist für ihn die Summe aus Subjekt-/Objektpositionen und deren Verhältnissen zueinander. Hierin liegt ein großer Unterschied zu vorangegangenen Raumverständnissen, da vorher Subjekte und Objekte meist singulär betrachtet wurden. Raumwahrnehmung war für Aristoteles z. B. entweder Wahrnehmung des eigenen Körpers oder

Beobachtung anderer Körper. Durch Leibniz' Raumverständnis wird dieser Betrachtungsweise eine zweite Dimension hinzugefügt: Der eigene Körper wird räumlich in Differenz zu anderen wahrgenommen.

Die Auseinandersetzung zwischen den beiden Gelehrten ist in diesem Zusammenhang von Bedeutung, da sie verdeutlicht, dass die Betrachtung des Raumes auf sehr unterschiedlichen Ebenen stattfinden kann. Der absolute, Newton'sche Raum ist ein Makroraum, in dem potentiell mehrere Welten Platz finden könnten (Clarke, 1717a, Absatz 21–25). Es ist ein apriorischer Raum, der völlig unabhängig von sämtlichen Lebensformen existiert. Das Leibniz'sche Raumverständnis ist zwar abstrakt, orientiert sich aber an konkreten, phänomenologischen Größen, da es durch diese bestimmt wird. Darum wird hier die Leibniz'sche Idee des Ordnungssystems Raum zumindest in ihren Grundzügen dem Newton'schen Modell des absoluten, unveränderlichen Makroraumes vorgezogen. Als konkrete, mess- und beobachtbare Größen, mit Hilfe derer Raum beschrieben werden kann, werden Subjekt- und Objektpositionen, sowie ihre Relationen untereinander[13] übernommen. Da hier eine ausgewogene Darstellung angestrebt wird, sei noch kurz angemerkt, dass auch das Newton'sche Raumverständnis dabei helfen kann, erzähltheoretische Raummodelle zu erstellen oder weiter zu verfeinern. Die weiter oben beschriebene Traditionslinie der Possible-Worlds-Theorie wird z. B. eher durch ein absolutes Raumverständnis gestützt.

Kant verwirft sowohl die Idee des absoluten als auch die des relativen Raumes, um sein Postulat des apriorischen Raumes, des Raumes als Quelle der Erkenntnis, zu entwickeln (Kant, 2006, 78). Der absolute Raum sei ein reines Gedankenkonstrukt und gehöre in die „Welt der Fabeln" (Kant, 2006, 78). Die Vertreter des relativen Raumes würden abstreiten, dass es einen solchen unabhängig von Objekten gäbe, und damit die Geometrie als abstrakte Wissenschaft anzweifeln (Kant, 2006, 78). Kant selbst nimmt die Position ein, dass der Raum eine schematische Denkfigur sei (Kant, 2006, 76 und 78), die einerseits der Erfahrung vorausgehe (Kant, 2006, 79 und Kant, 1993, 98) und andererseits Erkenntnis durch diese Erfahrung ermögliche (Kant, 2006, 82). Er verbindet also die Idee des absoluten Raumes als Denkfigur, die der Erkenntnis vorausgeht, mit der Wahrnehmung von Objekten, ihrer Lage, ihrer Individualität und ihren Relationen zueinander. Auf diese Weise legt er den Grundstein für die Idee des Anschauungsraumes:

[13] Eine Konkretisierung solcher Raumbeschreibungen anhand bestimmter Wortgruppen findet sich auch bei Dennerlein (2009). Bei der Operationalisierung wurde die Berücksichtigung der von Leibniz beschriebenen Raumelemente darum über Dennerleins Konkretisierung umgesetzt.

> Der Raum ist nichts anders, als nur die Form aller Erscheinungen äußerer Sinne, d.i. die subjektive Bedingung der Sinnlichkeit, unter der allein uns äußere Anschauung möglich ist. (Kant, 1993, 101)

Statt Raum also entweder in positiver (relativer Raum) oder negativer Weise (absoluter Raum) an Objekte zu binden, stellt Kant ihn in engen Zusammenhang mit Subjekten. Dies zieht auch den Umkehrschluss nach sich, dass Raum ohne die menschliche Wahrnehmung und Erkenntnis gar nichts sei (Kant, 1993, 102). Raum und Erkenntnis bedingen sich für Kant gegenseitig (Kant, 1993, 104).

Durch die enge Verknüpfung von Raum und Erkenntnis und damit äußerem Anschauungsraum und innerem Denkraum kommt bei Kant auch die Frage auf, ob es möglich sei, nicht-räumlich zu denken. Kant bejaht dies durch seine Erläuterungen *über die Zeit*. Der Zeitbegriff wird von Kant analog zum Raumbegriff apriorisch, abstrakt, schematisch und subjektiv gedacht. Während der Raumbegriff das Nebeneinander beinhalte, beschreibe der Zeitbegriff das Nacheinander (Kant, 1993, 109). Der große erkenntnistheoretische Unterschied ist, dass Zeit „die formale Bedingung a priori aller Erscheinungen überhaupt" (Kant, 1993, 109) sei. „Der Raum [...] ist als Bedingung a priori bloß auf äußere Erscheinungsformen eingeschränkt" (Kant, 1993, 109). Unter der Annahme, dass eine Erzählung nicht zwingend äußere Anschauung beinhalten muss, sondern auch auf innere Vorgänge – und damit auf Denkprozesse, die sich nicht zwingend auf räumliche Phänomene richten – beschränkt sein kann, lässt sich also schlussfolgern, dass es theoretisch möglich sein müsste, dass in einer Erzählung keinerlei Raumdarstellung evoziert wird. Das Kant'sche Raum- und Zeitverständnis stützt darum den konzeptionellen Rahmen dieser Untersuchung. Darüber hinaus ziehe ich für die Erzähltextanalyse aus Kants Ansätzen den Schluss, dass Raumdarstellung sowohl äußere Erscheinungsformen beschreiben kann als auch innere Vorgänge, die Gedanken über räumliche Aspekte umfassen. Anders als Schmid (2008, 24), der für die Definition des Ereignisbegriffes fordert, diesen auf tatsächlich in der storyworld stattfindende Zustandsveränderungen zu beschränken, nehme ich sämtliche erzählerisch dargestellten Räume in den Blick, seien diese nun durch die Erzählinstanz vermittelt oder von einer Figur erzählt oder bloß vorgestellt[14]. Es wird also im Folgenden nicht streng zwischen den Ebenen der mentalen Repräsentation in Erzähltexten und der Ereignisebene unterschieden. Zentral für die diese Studie ist lediglich die Frage, ob in der Darstellung Raumausdrücke und -hinweise enthalten sind.

---

[14] Zu dieser Perspektive trägt auch das Projektdesign dieser Studie bei. Primäres Anliegen ist es, zunächst einmal alle Referenzen räumlicher Informationen ausfindig machen und betrachten zu können, bevor im nächsten Schritt weiter ausdifferenziert und der Fokus auch wieder eingeschränkt werden kann.

Ein weiterer Ansatz, in dem Anschauung und somit die Beziehung zwischen Subjekt und Objekt ins Zentrum gerückt wird, ist der von Husserl in seiner Vorlesung zum Thema Ding und Raum (Husserl, 1991) dargelegte. Darin entwirft Husserl das Konzept einer Dingeinheit, die sich aus dem Objekt und der Anschauung des Objekts und seiner räumlichen Einbettung in die Umgebung zusammensetzt (Husserl, 1991, 234). Ausgangspunkt ist für ihn die Wahrnehmungsinstanz, die er als „Ich-Ding" (Husserl, 1991, 5) oder „Ichleib" bezeichnet (Husserl, 1991, 162). Die Besonderheit dieses wahrnehmenden Körpers gegenüber anderen sei, dass das Ich Empfindungen habe, die es in seinem Inneren lokalisiere (Husserl, 1991, 162). Außerdem gebe es für jede Wahrnehmung eine Motivation, ein Interesse, eine Meinung (Husserl, 1991, 129–130). Abgesehen von diesen Subjektkomponenten betrachtet Husserl die Wahrnehmungsinstanz sehr objektiv und ist eher an physikalischen Gegebenheiten (wie z. B. Position, Relation zum Objekt, Sichtfeld) interessiert als an psychologischen Vorgängen (vgl. Husserl, 1991, 161). Für diese Studie ist außerdem noch sein stark an die Ideen des euklidischen Raumes erinnerndes Verständnis des zweidimensionalen Feldes interessant. In einem solchen seien die kleinsten, möglicherweise unteilbaren Einheiten Punkte, die Punkte setzen sich zusammen zu linearen Begrenzungen, innerhalb derer sich das Feld als Kontinuität ergibt (vgl. Husserl, 1991, 166). Diese Kontinuität ist anders herum auch wieder in kleinere Einheiten teilbar, wahrscheinlich bis zum Punkt als kleinstem Element, möglicherweise aber auch bis ins Unendliche (vgl. Husserl, 1991, 166). Zentrale Termini, die Husserl für seine Analyse nutzt sind: Punkt, Linie, Lage, Figur, Größe (Husserl, 1991, 166).

Die ausgeführten Ansätze zum Anschauungsraum fügen der vorliegenden Analyse den grundlegenden Aspekt hinzu, dass die Wahrnehmungsinstanz räumlicher Ausgangs- und Bezugspunkt ist. Ganz konkret bedeutet das, dass in Ausdrücken wie „und da fiel ihr Blick auf eine Kiste" eine räumliche Achse aufgebaut wird, die zwischen einem Subjekt und einem Objekt verläuft. Darüber hinaus kann aus den Ausführungen Kants und Husserls geschlossen werden, dass der Körper der Wahrnehmungsinstanz ebenfalls ein wichtiger räumlicher Bezugspunkt ist. Kant nennt die Betrachtung der Hände als wichtiges Beispiel der Raumerfahrung und -erkenntnis (vgl. Kant, 2006, 74). Bei Husserl sind es vor allen Dingen Blick-Linien (vgl. Husserl, 1991, 42–59) und -Felder (vgl. Husserl, 1991, 82–84), Bewegungen zu Objekten hin oder von ihnen weg (wobei sowohl das Subjekt (hierfür ist der Begriff der Kinästhesie zentral (vgl. Husserl, 1991, 154–185)) als auch das Objekt (vgl. Husserl, 1991, 272–284) bewegt sein kann) und die Lokalisierung von Empfindungen im Körper (vgl. Husserl, 1991, 162), die den Raum konstituieren. Somit können Erwähnungen des gesamten Körpers oder von Körperteilen der Figuren in einem Erzähltext Hinweise auf den Raum enthalten. Selbstaussagen wie „er fühlte Trauer

in seinem Herzen“ sind nach Husserl eine Lokalisierung und können ebenfalls als Hinweise auf den Raum gedeutet werden.

Nahezu zeitgleich mit Husserls Ausführungen, nämlich im Jahre 1905, publizierte Einstein seine Ausführungen zu einer speziellen Relativitätstheorie. Dass Einsteins Betrachtungen von der Erzähltheorie, insbesondere in Bachtins *Chronotopos*, herangezogen wurden, ist in Abschnitt 2.2.2 bereits deutlich geworden. Die Hauptidee, die dort von Einstein übernommen wurde, ist die einer vierdimensionalen Raumzeit. Diese wird aufbauend auf Descartes und Minkowski als raumzeitliches Kontinuum verstanden (Einstein, 2006, 97–98). Einstein war als Vertreter des relationalen Raumes der Meinung, dass es keinen leeren Raum gäbe (ebenso wie Leibniz, s.o.). Ein solcher sei lediglich eine Denkfigur, die es leichter mache, räumliche Phänomene zu begreifen, da ein relativer und somit zumindest potentiell grenzenloser Raum zu komplex sei (Einstein, 2006, 94). Er nutzt in diesem Zusammenhang den Begriff der Mannigfaltigkeit, um die Komplexität des tatsächlichen (in seinen Augen relativen) Raumes zu beschreiben (Einstein, 2006, 94). Neben relativem, raumzeitlichem Kontinuum und Mannigfaltigkeit kann noch ein weiterer Faktor aus Einsteins Theorie auf die Erzähltextanalyse übertragen werden: Das Äquivalenzprinzip. Demnach sind bestimmte Eigenschaften von Körpern unter bestimmten Bedingungen ununterscheidbar. Ob eine Masse schwer oder träge ist, kann auf lokaler Ebene nicht unterschieden werden. Ob eine Masse schwerelos ist oder sich im freien Fall befindet, kann ebenfalls nicht entschieden werden, wenn man nur ein ganz kleines raumzeitliches Gebiet betrachtet (zum Äquivalenzprinzip vgl. Schmidt, 2009, 302). In literarischen Texten werden meistens nur mehr oder weniger kleine Ausschnitte einer narrativen Welt oder gar eines narrativen Universums thematisiert (vgl. z. B. Lotman, 1973, 300–305; Bauer, 1997, 74–75). Raumdarstellungen in narrativen Texten sind immer lückenhaft (vgl. Stanzel, 2008, 157). Nach dem Äquivalenzprinzip kann also nicht abschließend darüber geurteilt werden, wie der gesamte Raum eines literarischen Narrativs beschaffen ist. Anders ausgedrückt: Rückschlüsse von Raumeinheiten und Raumkomplexen, die im Text angelegt sind, auf eine allumfassende Topologie einer storyworld, so wie sie in Abschnitt 2.2.3 in Bezug auf Zoran bereits als problematisch für diese textzentrierte Analyse beschrieben wurde, sind in der Tat nicht möglich. Hier wird die Entscheidung, diese Ebene aus der Betrachtung auszuschließen erneut – wenn auch aus ganz anderer Perspektive – bestätigt.

Dass es durchaus sinnvoll ist, Grundideen von Einsteins Äquivalenzprinzip hier in die Analyse literarischer Räume einzubeziehen, zeigen zwei kurze Beispiele, die Dennerlein in ihrer *Raumnarratologie* einbringt. Dabei geht es darum, dass Objekte in Erzähltexten Qualitäten annehmen können, die ansonsten eher Orten vorbehalten sind. Das betrifft vor allem die Positionierung von Figuren. Dennerlein nennt

als Beispiele für Dinge, in denen sich Figuren aufhalten können die Streichholzschachtel und die Flasche (vgl. Dennerlein, 2009, 77). Außerdem können abstrakte Größen in narrativen Texten örtlich werden. Vor allem in filmischen Erzählungen wird z. B. Zeit häufig als betretbarer Ort dargestellt, wenn Figuren sich auf eine Zeitreise begeben[15]. Nach dem Äquivalenzprinzip entsteht hier das Problem, dass auf der Satzebene nicht immer eindeutig entschieden werden kann, ob etwas ein Ort oder ein Objekt ist – der betrachtete Kontext ist dafür zu klein. In dem Satz „Der Geist ist in seiner Flasche" könnte „Flasche" sowohl ein Ort (geographisch fest positioniert) als auch ein Objekt (nicht an eine geographische Position gebunden) und damit lediglich ein Raumhinweis sein. Gibt es nicht genügend weitere Kontextinformationen, die die Kategorisierung ermöglichen, so wird es für einen Computer unentscheidbar, in welche Kategorie das Wort „Flasche" hier fällt. Erst ein zweiter Satz wie „Er nahm die Flasche samt Geist und ging damit fort.", eine Öffnung des betrachteten Kontextes, würde eine Entscheidung über die Zugehörigkeit zu einer bestimmten Kategorie zu Gunsten des Objektcharakters, ermöglichen (Die Textbeispiele wurden von mir konstruiert, bei Dennerlein finden sich diesbezüglich keine konkreten Satzbeispiele). Dabei bleibt aber die inhärente Ambivalenz (tatsächlich hat die Flasche hier wohl sowohl Qualitäten eines Ortes als auch die eines Objektes, das neben Hinweisen auf den Raum noch einen ganzen Assoziationskomplex kommuniziert). Beim Beispiel der Zeitreisen eröffnet sich noch eine andere Problematik. Der Fokus auf ein Figuren-Ensemble macht es unmöglich zu entscheiden, ob Zeit in der gesamten storyworld (topologische Ebene nach Zoran) in mehr als einem Ausnahmefall zu einem Raum wird, der betreten werden kann, ob es sich also um eine allgemeine Gesetzmäßigkeit handelt oder einen Bruch mit den Gesetzmäßigkeiten der narrativen Welt. Da in dieser Studie keine Genre-Grenzen berücksichtigt werden, findet hier das Principle of Minimal Departure (Ryan, 2012b, Absatz 6) Anwendung. Dieses von Ryan eingeführte Prinzip besagt, dass die erzählte Welt so lange als der Welt der Lesenden gleich angenommen wird, bis eine Abweichung explizit dargestellt wird. Das heißt konkret, dass hier z. B. ein geographischer Ortsbegriff angenommen wird. Objekte werden nur dann zu Orten, wenn sie im Text als solche dargestellt werden. Da das eingesetzte Machine-Learning-Tool kontextsensitiv

[15] Sehr prominent ist diese Zeit-als-Raum-Darstellung z. B. in der Superhelden-Serie *The Flash*, in der der Protagonist so schnell laufen kann, dass sich ein rötlicher Lichtstrahl bildet, auf dem er sowohl in die Zukunft als auch in die Vergangenheit rennen kann. Sowohl Zeit als auch Licht werden hier als manifester Raum dargestellt; eine genretypische Referenz auf wissenschaftliche Ideen und Entdeckungen wie die Lichtgeschwindigkeit und das bereits erwähnte Einstein'sche Raum-Zeit-Kontinuum. Ein weiteres Beispiel liefert der Film *Avengers – Endgame*, in dem sich für die Protagonisten bei einer Zeitreise eine Art farbenfroher senkrechter Tunnel öffnet, den sie nutzen, um in die Vergangenheit zu springen.

ist, können explizite Abweichungen von der Referenzwelt durchaus Berücksichtigung finden. Befinden sich diesbezügliche Informationen allerdings außerhalb des Kontext-Fensters, das in der Kombination betrachteter Wort- und Satzmerkmale angelegt ist, so ist es für die Software ununterscheidbar, ob etwas in eine Raumkategorie fällt oder in eine andere. Hier greift also das Äquivalenzprinzip.

2.3.3 Kultureller und sozialer Raum

Ausgangspunkt einiger zentraler Kulturtheorien zum Raum ist die Differenzierung zwischen Natur und Kultur und damit zwischen zwei Raumqualitäten. Der Naturraum ist zwar Ausgangspunkt und Basis solcher Kulturtheorien, er kann aber nicht in seiner Ursprünglichkeit betrachtet werden (vgl. dazu z. B. Lefebvre, 2006, 330). Der Mensch ist als soziales, denkendes und erlebendes Wesen nicht fähig, den Naturraum vor dem Eingreifen der Menschheit zu durchdenken (vgl. Lefebvre, 2006, 330).

Angesichts der Tatsache, dass bereits im 18. Jahrhundert aus gesellschaftstheoretischer Perspektive von Rousseau ein Akt der Herrschaft über den (Natur-)Raum beschrieben wurde[16], um die Entwicklung der bürgerlichen Gesellschaft zu beschreiben, mag es verwundern, dass in den Kulturwissenschaften erst etwa ab dem Jahr 1967 ein spatial turn einsetzte (Döring, 2008, 9), der erst im Jahre 1989 auch als solcher benannt wurde (Döring, 2008, 7). Doch insgesamt lässt sich hier wohl ein ähnliches Bild zeichnen wie in der literaturwissenschaftlichen Forschung zum Thema Raum: Obwohl Raum schon lange als Bestandteil von Kultur betrachtet wurde, setzte eine intensive Betrachtung kultureller Räume erst in der zweiten Hälfte des 20. Jahrhunderts ein. Häufig wird in der kulturwissenschaftlichen Forschung Kultur als additive Dimension betrachtet, die entweder Orte zu Räumen machen kann (so z. B. in Lefebvre, 2006; Deleuze, 2006; Certeau, 2006) oder bestehende Räume mit weiterer Bedeutung auflädt (z. B. in Cassirer, 2006; Bourdieu, 2006). Orte und Räume werden in den Kulturwissenschaften sowohl mit Blick auf die Konstitution des Raumes als solchem, also topologisch, als auch mit Fokus auf die semiotische Aufladung, also topographisch, betrachtet (vgl. Günzel,

[16] Rousseau schreibt in seiner 1755 erschienen Abhandlung über die Entstehung der Ungleichheit:

> Der erste, der ein Stück Land eingezäunt hatte und dreist sagte: ‚Das ist mein' und so einfältige Leute fand, die das glaubten, wurde zum wahren Gründer der bürgerlichen Gesellschaft. (Rousseau, 2017, 191)

2008). Nach dem topographischen Raumverständnis sind Orte nicht per se kulturwissenschaftlich interessant, sondern nur in ihrer kulturellen Prägung. Im Hintergrund steht ein Topographie-Begriff, der sich auf das buchstäbliche Einschreiben von Bedeutungen in den Raum bezieht und der sehr nah am Kultur-als-Text-Paradigma angesiedelt ist (vgl. Döring, 2008, 17). Steht der topologische Ansatz im Vordergrund, so sind mathematische Grundannahmen zum Raum meist wichtige Orientierungspunkte (Günzel, 2008, 224–225). Kulturwissenschaftliche Raumforschung wird dann besonders fruchtbar, wenn sowohl topologische als auch topographische Raumaspekte betrachtet werden und wenn diese Differenzierung klar berücksichtigt wird (Günzel, 2009a, 230). Dies ist zum Beispiel bei Deleuze und Guattari erkennbar, die zwischen glattem Raum, einer Entsprechung zum topologischen Raum, und gekerbtem, topographischen Raum, unterscheiden. Der glatte Raum ist hier stets ein unberührter Raum. Längen, Verbindungen und Flächen sind von besonderer Bedeutung. Betrachte man hingegen den gekerbten, vom Menschen mit Einschnitten versehenen Raum, so gehe man von einem Punkt zum nächsten (Deleuze, 2006, 434–436).

Bisher wurde in dieser Arbeit, ausgehend von Descartes, die grundlegende Unterscheidung zwischen Orten als topographischen Referenzpunkten und Raum als multidimensionalem Konzept, welches physikalische und semiotische Aspekte gleichzeitig enthalten kann, als definitorische Voraussetzung angenommen. Der Vollständigkeit halber muss aber noch erwähnt werden, dass es auch andere Begriffsdefinitionen des Begriffspaares Ort-Raum gibt. In den Kulturwissenschaften schlägt z. B. Assmann vor, Raum als ein Konzept der aktuellen und zukunftsgerichteten Gestaltung anzusehen. Orte seien ihrem Verständnis nach von der Vergangenheit geprägt (Assmann, 2015, 16). Dieses Verständnis wird manchmal nahezu als das Gegenteil der Definitionen verstanden, die dieser Arbeit zu Grunde liegen. So fassen z. B. Csáky und Leitgeb zusammen, dass ein „bedeutungsneutraler, physischer ‚Raum' zu einem kulturell definierten und beladenen ‚Ort"' werden kann (Csáky und Leitgeb, 2009, 7). Obwohl Assmann weniger eine solche starre Dichotomie beschreibt, sondern eher einen Unterschied der Perspektive auf denselben Gegenstand – nämlich in seiner Gemachtheit oder in seiner Prägbarkeit (Assmann, 2015, 22) – ist es an dieser Stelle wichtig festzuhalten, dass es Ansätze gibt, in denen die Ort-Raum-Differenzierung grundlegend anders definiert ist als hier. Diese Tatsache muss bei der nun folgenden Betrachtung kulturwissenschaftlicher Raummodelle berücksichtigt werden.

Kulturwissenschaftliche Ansätze zur Erforschung des Raumes sind divers und konzentrieren sich häufig auf die Bedeutungsebene des Raumes. Im Folgenden sollen darum zwei Ansätze herausgegriffen werden, die den Raum hauptsächlich in seiner Struktur und / oder elementaren Beschaffenheit betrachten. Die berücksich-

tigten Raumkonzepte sind der ästhetisch-kulturelle Raum und der soziale Raum. Als ergänzender Diskurs wird abschließend der Raumbegriff der Archäologie betrachtet, der eine neue Perspektive auf die Verbundenheit von Raum und Zeit ermöglicht.

Der ästhetisch-kulturelle Raum

Der Feststellung, dass es einen ästhetischen Raum oder vielmehr ein ästhetisches Raumverständnis gibt, liegt die Annahme zu Grunde, dass der Raum verschiedene Qualitäten annehmen kann. Bereits 1931 differenzierte Cassirer zwischen mythischem, ästhetischem und theoretischem Raum (Cassirer, 2006, passim). Ähnlich wie andere Denker vor ihm (namentlich Hegel und Heidegger, s. Abschnitt 2.3.1 stellt auch Cassirer die Frage nach dem ontologischen Status von Raum (Cassirer, 2006, 488). Er überwindet die Frage nach dem Sein des Raumes allerdings, indem er dem Raum statt des Seins, welches die Objekte kennzeichnet (Cassirer, 2006, 488), die Ordnung und damit einen grundsätzlichen Pluralismus zuweist (Cassirer, 2006, 492). Raum ist nach Cassirer keine feststehende Ordnung, sondern kann nach verschiedenen Prämissen, die alle der Sinnstiftung dienen, geformt werden. Dazu gehört der mythische Raum ebenso wie der theoretische und der ästhetische Raum (Cassirer, 2006, 494). „Die Sinnfunktion ist das primäre und bestimmende, die Raumstruktur das sekundäre und das abhängige Moment“, stellt Cassirer fest und zeigt damit auf, dass Raum keine feste Form hat, sondern je nach Qualität seiner Ordnung eine individuelle Prägung erhält (Cassirer, 2006, 494). Nach diesem Raumverständnis ist die Geometrie nur eine Ausprägung, die das Sinn- und Ordnungssystem annehmen kann. Ein anderes Raumverständnis kann jede Richtung, jede Lage, jeden Punkt und jede Verbindung mit einer vom Raumverständnis abhängigen Bedeutung belegen (Cassirer, 2006, 495). Hier sind wir also bereits sehr nah an dem semiotischen Raumverständnis von Bachelard und Lotman (vgl. Abschnitt 2.2.1). Der ästhetische Raum ist nach Cassirer (2006, 497) eine reflexive Ordnung. Denn in der Nachahmung in den Künsten setzt sich der Mensch in ein Verhältnis zur Welt (und damit ist ästhetisches Schaffen niemals bloße Nachahmung). Trotzdem ist der ästhetische Raum, anders als der theoretische, kein hauptsächlich vom Denken geprägter Raum. Es ist ein Raum der Emotion und der Phantasie (Cassirer, 2006, 498). Im Gegensatz zum mythischen Raum ist der ästhetische Raum ein vom Menschen beherrschter, gestaltbarer Raum (Cassirer, 2006, 499). Der ästhetische Raum entspricht damit aber nicht dem gekerbten Raum von Deleuze und Guattari (s.o.). Da der ästhetische Raum eine reflexive Ebene besitzt, ist es möglich, sowohl glatte Räume als auch gekerbte Räume und mehr noch glatte und gekerbte Raumerfahrungen im ästhetischen Raum auszudrücken (Deleuze, 2006, 442–443).

Ein weiteres ästhetisch-kulturelles Raumkonzept ist das des Bewegungsraums. Diesem Konzept liegt die Idee zu Grunde, dass Räume, im Gegensatz zu Orten

durch eine Form der Bewegung erschaffen werden (Certeau, 2006, 346). In Alltagserzählungen wird Raum häufig als Wegstrecke und in Form von Anweisungen (wie bspw. „du wendest dich rechts und dann bist du im Garten“) konstruiert (Certeau, 2006, 347–348). Sehr viel seltener sind Darstellungen als eine Art Karte, in der Orte so beschrieben werden, als sehe man sie auf einem Plan (ebd.). Über den „Umweg“ über den Film als Medium audio-visueller Narration wird deutlich, dass Raum und Bewegung gemeinsam zur Konstruktion von Geschichten beitragen (Certeau, 2006, 346). In der konstruierten Szenerie eines Films werden vor stillstehender Kamera schon kleinste Bewegungen im Raum, wie z. B. das Krümmen eines Fingers (Rohmer, 2006, 515) zu Bedeutungsträgern. Der Blick auf diese minimalen Bewegungen zeigt, wie essentiell Raum für ein Narrativ ist, da Bewegung ohne Raum nicht möglich wäre und damit auch keine Handlung einsetzen könnte. Die Idee der minimalen Bewegungen ist sehr gut anschlussfähig an die oben beschriebenen Betrachtungen von Husserl zur Dingeinheit. Denn auch bei Husserl können es Minimalbewegungen (z. B. Augenbewegungen Husserl, 1991, 42–59) sein, die die Perspektive auf ein Objekt verändern oder sogar ein anderes / weiteres Objekt ins Blickfeld rücken. Die Ideen zum Bewegungsraum mit Husserls Idee der Dingeinheit verbindend, werden hier Bewegungen als Hinweise auf den Raum gedeutet, wenn sie zur Darstellung einer Relation zwischen zwei Dingen (Subjekten und / oder Objekten) beitragen.

Der soziale Raum
Dem sozialen Raum liegt ein relationales Raumverständnis zu Grunde. Der Raum wird durch soziales Handeln und Interaktion erzeugt (Döring, 2008, 25). Da diese Form der Raumerzeugung gesellschaftlich vonstatten geht, ist sie nicht unbedingt eine bewusste Entscheidung einzelner (Döring, 2008, 26). Simmel (2006) leitet sehr anschaulich her, inwiefern menschliches Leben und Räumlichkeit zusammenhängen. Anhand der Herrschaft über Territorien, die gleichzeitig eine Herrschaft über Menschen ist, macht er deutlich, dass es nahezu unentscheidbar ist, ob über den Raum Menschen oder über Menschen der Raum regiert wird (Simmel, 2006, 304). Anhand dieses einfachen, machtpolitischen Beispiels zeigt er, dass Mensch und Raum unmittelbar miteinander in Beziehung stehen. Der Mensch organisiert sein Sozialleben räumlich. Das geht von familiären Einheiten (das Haus) über soziale Gruppen (der Club, die Universität) bis hin zu Staaten oder staatenähnlichen Zusammenschlüssen (wie z. B. indigene Volksstämme) (Simmel, 2006, 307). Aber nicht nur besetzte Gebiete sind definiert, auch die neutralen Zonen dazwischen haben als solche eine soziale Bedeutung (Simmel, 2006, 310). Innerhalb der menschlichen Territorien gibt es ebenfalls eine stark raumbeherrschende Organisation (Simmel, 2006, 314–315); Räume der Ruhe und der Bewegung sind definiert, Räume des Lernens und der Freizeit ebenfalls und es gibt noch unzählige weitere Beispiele

räumlich-sozialer Organisation. Mit seinen Ideen zum sozialen Raum legt Simmel einen Grundstein für spätere sozio-kulturelle Raummodelle. So ist hier z. B. bereits der Zusammenhang von Räumen, gesellschaftlichen Funktionen und Macht angelegt, der später für Foucaults diskursiv geprägtes Raummodell von Bedeutung ist. Und auch die Ansätze Bourdieus zum sozialen Raum lassen sich als Ergänzung dessen, was Simmel beschreibt, um weitere Ebenen auslegen.

Dass der Raum nicht leer ist, sondern immer mit unterschiedlichen Qualitäten belegt und dass Raum (heute) relational gedacht wird, sind die zentralen Ideen von Foucaults Raumverständnis (vgl. Foucault, 2006, passim). Dabei unterscheidet Foucault binär zwischen sozialem Raum und individuellem Raum (Foucault, 2006, 319) und in einem dreiteiligen Modell zwischen „normalem" Raum (der bei ihm undefiniert bleibt), Utopie, die unverwirklicht bleibt und Heterotopie, die die verwirklichte Utopie ist. Nach Foulcaults Heterotopologie entwickelt jede Gesellschaft Räume für Phänomene, die als andersartig, als Abweichung von Normalität, empfunden werden (Foucault, 2006, 321–322). Die Art der Heterotopien ist abhängig von der Gesellschaft, die sie hervor bringt (Foucault, 2006, 322–327). Nach Foucault gibt es folgende Heterotopien: Jugend-, Alten- und Erholungsheime, Psychiatrische Anstalten, Gefängnisse, Kasernen, Friedhöfe, Initiationsräume und andere kultische Stätten, Theater, Gärten, Museen, Bibliotheken, Festwiesen, Feriendörfer und Hotels, Bordelle, Kolonien und Schiffe (Foucault, 2005, 11–22). Spiegel sind eine interessante Zwischenform von Utopie und Heterotopie, da der Ort, den sie zeigen zwar utopisch ist, gleichzeitig zeigen sie aber auch, was da ist, nämlich den davor stehenden Körper, den sie durch diese Manifestation von allem Utopischen befreien (Foucault, 2005, 35–36). Der Spiegel ist damit ein besonders interessantes Raumsymbol. Hinter diesem Ansatz steht ebenfalls die Idee der sozialen Organisation räumlicher Gegebenheiten. Diese Idee folgt dem Grundsatz, dass Kulturen sich mit Räumen verknüpfen, indem sie Räume schaffen, die bestimmte individuelle und kollektive Erfahrungen organisieren. Der Raum wird hier zum kulturellen Organisationsprinzip, das in unmittelbarer Abhängigkeit zum Diskurs steht. Denn bei jeder qualitativen Aufladung von Raum ist zunächst die Frage entscheidend, ob normatives Verhalten organisiert werden soll oder abweichendes (wobei die Abweichung bei Foucault unbewertet bleibt, ein Garten ist bei Foucault ebenso eine Heterotopie wie ein Gefängnis). Der alltägliche Raum und die Heterotopien funktionieren also analog zum Diskurs (wie Foucault ihn in *Die Ordnung der Dinge* und *Archäologie des Wissens* definiert) und Gegendiskurs (Foucault, 2008, 80); letztendlich geht es immer um die gesamtgesellschaftliche Suche nach einem Rahmen für menschliches Erleben und Verhalten.

Auch Bourdieus (Bourdieu, 2006) Aussagen zum Raum folgen dem Grundsatz, dass gesellschaftliche Organisation mit Lebensräumen verknüpft wird. Für ihn ist

der soziale Raum ein komplexes, multidimensionales System, welches Orte mit Tätigkeiten, Tätigkeiten mit Berufsgruppen und Berufsgruppen und Tätigkeiten mit sozialem Status (genauer: kulturellem und ökonomischem Kapital Bourdieu, 2006, 358) verbindet. Zum Beispiel sind Tätigkeiten in der Natur wie Bergsteigen oder Wandern mit einem sozialen Milieu verknüpft, welches über ein hohes kulturelles Kapital und ein geringes ökonomisches Kapital verfügt (Bourdieu, 2006, 357). Bourdieus System erinnert durch die systematischen Zuschreibungen an Lotmans semiotischen Raum. Doch weist es sehr viel mehr Dimensionen auf und verbindet multiple räumliche Strukturen miteinander. Topologischer Raum und Raumzeit sind für Bourdieu durchaus von Bedeutung, denn er erarbeitet seine Parameter (kulturelles und ökonomisches Kapital, Berufsgruppen und Tätigkeiten) am Beispiel von Frankreich in den 1970er Jahren (Bourdieu, 2006, 354). Bei Bourdieu sind mehrere Raummodelle erkennbar. Der topologische oder ortsgebundene Raum wurde bereits erwähnt. Dazu kommt ein skalares und ein vektorielles Raumverständnis. Anhand zweier Vektoren misst er kulturelles und ökonomisches Kapital und verbindet die beiden zu einem Gesamtkapital, das positiv oder negativ gerichtet sein kann. Das skalare Raumverständnis zeigt sich z. B. wenn er Musikinstrumente anhand dieser Vektoren ausrichtet. Während er das Akkordeon z. B. beim niedrigen Gesamtkapital verortet (etwas überdurchschnittliches ökonomisches Kapital, geringes kulturelles Kapital), gesteht er dem Klavier ein sehr hohes Gesamtkapital zu (hohes kulturelles Kapital, mittleres ökonomisches Kapital) (Bourdieu, 2006, 357). Außerdem gibt es noch ein systemisches Raumverständnis. Akteure und ihre Dispositionen (ihr Habitus) gestalten ein relationales räumliches System, in dem vor allem die Differenzen zwischen den sozialen Positionen für die Verortung entscheidend sind (Bourdieu, 2006, 359).

> In Wirklichkeit ist der zentrale Gedanke, daß in einem Raum existieren, ein Punkt, ein Individuum in einem Raum sein, heißt, sich unterscheiden; unterschiedlich sein (Bourdieu, 2006, 361).

Dieses System wirkt dann auch auf den topologischen Raum zurück (wie am Beispiel von Paris gezeigt).

Die hier angeführten soziologisch-kulturwissenschaftlichen Ansätze zur genaueren Ausdifferenzierung des Raums haben dem Raummodell, das hier entwickelt wurde, auf zwei Ebenen Aspekte hinzugefügt. Zum Einen muss auf Ebene der Raumthemen die kulturelle und soziale Aufladung von Räumen mit berücksichtigt werden. Das Raumthema Reise kann z. B. auf unterschiedliche Weise kulturell aufgeladen sein. Heterotopien werden als Raumthemen integriert, die kulturell von besonderer Bedeutung sind. Zum Anderen kann die Ebene der Hinweise auf den

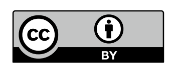
Raum nach den kulturwissenschaftlichen Betrachtungen genauer beschrieben werden. Bewegungen sind wichtige Hinweise auf Raum. Soziales Handeln, Macht und / oder Territorien verweisen auf räumliche Strukturen oder stehen mit solchen in enger Verbindung. Auch die gesellschaftliche Struktur spiegelt sich im Raum wider. Dabei ist die Verknüpfung tatsächlicher sozialer Strukturen mit der Literatur ebenfalls von Bedeutung. Die Pariser Gesellschaft, die sich auch in der Topographie der Stadt zeigt, wie sie von Bourdieu beschrieben wurde, kann zur Grundlage literarischer Werke werden. An dieser Stelle soll nun als letztes ein kurzer Exkurs gewagt werden, um das Raummodell noch ein Stück weiter theoriegeleitet zu begründen. Dieser Exkurs führt in die Archäologie, die Raum aufgrund ihres Forschungsgegenstandes ganz anders betrachten kann.

2.3.4 Exkurs: Raum in der Archäologie

Es ist weder möglich in dieser Arbeit, die komplette Literatur zum Thema Raum in der Archäologie zu beschreiben, noch Ziel, einen Beitrag zur Raumforschung aus dieser Perspektive zu leisten. Aber einige zentrale Ansätze und Techniken dieser Disziplin helfen, genauer zu verstehen, welche Qualität der Raum hat, da dieser nicht hauptsächlich oberirdisch und damit in der Luft als nahezu nicht-haptischem Element betrachtet wird. Stattdessen wird Raum sowohl oberirdisch als auch unterirdisch angeschaut und damit auch in dem festen Element der Erde gesehen. Mit Hilfe eines Gedankenexperiments, bei dem man Objekte archäologisch, also zunächst einmal in der Erde, denkt, lässt sich Descartes Ansatz des ausgefüllten Raumes viel besser nachvollziehen, denn in der Erde sind tatsächlich alle Zwischenräume gefüllt. Darüber hinaus stehen die einzelnen Erdschichten für Zeitschichten, d. h. nicht nur Raum, sondern auch Zeit werden in der Erde manifest. Bei der archäologischen Arbeit werden Erdschichten nach und nach abgetragen und die Objekte wieder in die Nachbarschaft der Luft überführt. Gleichzeitig wird anhand der Erdschichten abgeschätzt, seit wann die Objekte jeweils unter der Erde sind. Schließlich wird anhand ihrer Lage zu rekonstruieren versucht, welche überirdischen Positionen sie einmal eingenommen haben. (Lang, 2009, 30–31) In der archäologischen Forschung werden sowohl soziale als auch kulturelle Bedingtheit des Raumes mitgedacht. Am Beispiel des Landschaftsraums, der überirdisch betrachtet wie ein Naturraum wirken kann, zeigt sich besonders deutlich, dass die Trennung zwischen Natur und Kultur, die in den Kultur-

wissenschaften häufig versucht wurde, nicht haltbar ist. Denn unter einer Landschaft verbergen sich häufig ältere, kulturell aufgeladene Schichten, die lediglich oberflächlich überdeckt wurden. Natur und Kultur befinden sich in einem Wechselspiel und sind nicht voneinander trennbar (Lang, 2009, 32). Die enge Verknüpfung von räumlichem Zusammenhang (Fundort) und kultureller Funktion wird in der Archäologie besonders deutlich. Zwei vergleichbare Gegenstände können unterschiedliche Bedeutungen annehmen, wenn einer z. B. in einem Heiligtum gefunden wurde (religiöses Symbol), der andere aber in einer Villa (Symbol des Luxus) (Lang, 2009, 33).

Für diese Studie ist vor allem die enge Verbindung zwischen Kultur und Natur relevant. Bisher scheint es, als würde eine kulturelle Aufladung nicht unbedingt Positionen, also Orte und Relationen, betreffen. Durch Lotman wurde bereits klar, dass Relationen zwar eine binäre Aufladung haben können, die kulturell bedingt ist. Die archäologische Perspektive zeigt aber, dass das noch weitergedacht werden kann. Jeder Aspekt des Raumes kann bereits eine komplexe kulturelle Bedeutung transportieren. Mehr noch: Das Verhältnis einzelner Raumobjekte zueinander hat stets eine individuelle kulturräumliche Bedeutung. Damit wird Raum zu einem Werkzeug der Interpretation. Dieser Aspekt wurde in das operationalisierbare Raummodell mit aufgenommen.





3 Vermessung narrativer Raumdarstellung

Die Betrachtung des Raumes aus der Perspektive unterschiedlicher Disziplinen hat gezeigt, dass mit einer weiten Definition des Raumbegriffes wie z. B. die der res extensa von Descartes, ein Erzähltext ohne Raumdarstellung nicht denkbar wäre. Es hat sich aber auch gezeigt, dass Raum mit sprachlichen Ausdrücken mehr oder weniger explizit referenziert werden kann. Daraus schließe ich, dass Raum in Erzähltexten ein skalar ausgeprägtes, mehrdimensionales Phänomen sein muss. Analog zum Raum als mehrdimensionalem System liegt dieser Arbeit ein Verständnis von Erzählungen zu Grunde, nach dem es neben der Linearität der fortlaufenden Handlung auch versteckte Muster und mehrere in einzelne Einheiten unterteilbare Dimensionen gibt, die analytisch herausgearbeitet werden können. Dieses Verständnis von Narrativen wird in Abschnitt 3.1 definitorisch begründet. Daran anschließend werden in Abschnitt 3.2 die Definitionen zu Erzähltexten mit den narratologischen Betrachtungen zur Darstellung von Raum aus dem Abschnitt 2.2 zusammengeführt. Im Fokus dieser Studie steht die narrative Darstellung von Raum, sodass sowohl in Abschnitt 3.1 als auch 3.2 vor allem narratologische Ansätze herangezogen werden, die Raumdarstellungen als eine zentrale Facette narrativer Texte beschreiben.

3.1 Definitorische Grundlagen des „Erzähltextes“

Bisher wurde Raum in seiner Bedeutung für literarische Texte und auch als phänomenologische Größe betrachtet, ohne dass der Gegenstand dieser Untersuchung klar definiert wurde. An dieser Stelle wird das nachgeholt, indem eine Arbeitsdefinition gefunden wird, die festhält, wie Erzählungen und Erzähltexte im Rahmen dieser Studie verstanden werden. Diese Arbeitsdefinition ist zugespitzt auf die Untersuchung literarischen Raumes und fußt hauptsächlich auf narratologischen Ansätzen, die Raumdarstellungen als zentralen Aspekt von Erzähltexten berücksichtigen. Genette

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_3

hat darauf hingewiesen, dass unter dem Begriff „Erzählung" mindestens drei Dinge verstanden werden können:

> In einem ersten Sinn – der heute im gewöhnlichen Gebrauch am evidentesten und zentralsten ist – bezeichnet *Erzählung* [Hervorhebungen im Original Anm. MKS] die narrative Aussage, den mündlichen oder schriftlichen Diskurs, der von einem Ereignis oder einer Reihe von Ereignissen berichtet [...]. In einem zweiten, weniger gebräuchlichen Sinn [...] bezeichnet *Erzählung* die Abfolge der realen oder fiktiven Ereignisse, die den Gegenstand dieser Rede ausmachen, und ihre unterschiedlichen Beziehungen zueinander – solche des Zusammenhangs, des Gegensatzes, der Wiederholung usw. [...] In einem dritten Sinn, der wahrscheinlich der älteste ist, bezeichnet *Erzählung* noch ein anderes Ereignis: diesmal nicht mehr das, von dem erzählt wird, sondern das, das darin besteht, daß jemand etwas erzählt: den Akt der Narration selber. (Genette und Knop, 1998, 15)

Genette legt den Fokus seiner Betrachtungen zwar auf das in seiner Definition erstgenannte Verständnis von Erzählung als narrativer Aussage, wie er im Weiteren ausführt, kann dieses Verständnis aber nicht vollkommen losgelöst von den anderen betrachtet werden (Genette und Knop, 1998, 16). Er schlägt darum vor, unter dem Begriff „Geschichte" (histoire) die Gesamtheit der erzählten Ereignisse zu verstehen, unter „Erzählung" (récit) den Diskurs, der davon erzählt und unter „Narration" (narration) den Akt des Erzählens (Genette und Knop, 1998, 199). Chatman definiert „Erzählung" grundsätzlich als semiotische Struktur. Mit dem der Genette'schen histoire ähnlichen Begriff „story" benennt er den Inhalt einer Erzählung und unter „discourse" versteht er Ausdruck und Vermittlung der Erzählung (Chatman, 1993, 19), die unabhängig vom Medium eine gewisse Struktur aufweisen (Chatman, 1993, 22–26).

Zur Abgrenzung einer Erzählung gegenüber anderen Formen der Kommunikation und kommunikativen Texte kann z. B. die Beschreibung dienen (Schmid, 2008, 22; Herman, 2007, 9; Ryan, 2007, 26–27; Abbott, 2011, §9 und §12). Nun sind aber die Kategorien Erzählung und Beschreibung nicht ganz trennscharf, sondern können ineinander übergehen (Herman, 2007, 8–9); Beschreibungen können narrative Elemente beinhalten (Herman, 2007, 8–9) und Erzählungen sind passagenweise beschreibend (Schmid, 2008, 21). Um zum Kern dessen vorzudringen, was sie ausmacht, können zentrale Qualitäten prototypischer Erzählungen betrachtet werden. Herman schlägt vor, dass prototypische Erzählungen folgende Aspekte repräsentieren sollten:

> (i) a structured time-course of particularized events which introduces
> (ii) disruption or disequilibrum into storytellers' and interpreters' mental model of the world evoked by the narrative (wether that world is presented as actual, imagined, dreamed, etc., conveying

(iii) what it’s like to live through that disruption, that is the „qualia“ (or felt, subjective awareness) of real or imagined disruptive experiences (Herman, 2007, 9).

Zwei Dinge sind an Hermans Vorschlag für diese Studie besonders relevant. Zunächst einmal die Erwähnung des Weltbegriffes, der, wenn auch nicht ausschließlich, eine räumliche Dimension dessen, was erzählt wird, impliziert. Neben Zeit und Figuren ist die Darstellung von Raum für Herman eines der drei Basiselemente von Erzählungen (Herman, 2012, 17). Das Erzählen ist für ihn ein Schaffensakt mentaler Welten, die von Rezipierenden imaginativ bewohnt und interpretativ mitgestaltet werden können (Herman, 2012, 14–15). Eine solche storyworld kann als manifest (innerhalb der storyworld), vorgestellt, geträumt oder anderweitig repräsentiert dargestellt werden, entscheidend ist lediglich eine subjektive Erfahrung bzw. Erfahrbarkeit, die von erzählender und interpretierender Instanz geteilt wird. Dass menschliche Erfahrung eine, oder sogar die, zentrale Basis von Erzählungen ist, zeigt auch Fludernik (Fludernik, 2011, passim) und geht dabei in ihrer Bewertung dieses Faktors sogar noch über Herman hinaus (Abbott, 2011, §30). Ebenso wie im dieser Studie zu Grunde liegenden Raumbegriff steht also auch bei der Erzählung menschliche Erfahrung im Zentrum.

Auch Ryan erwähnt in ihrer „fuzzy-set“-Definition (Ryan, 2007, 28) von Narrativen die Darstellung von Raum als eine notwendige Dimension (Ryan, 2007, 29). Weitere, zum Teil mit mehreren Unteraspekten versehene, Dimensionen sind die zeitliche, mentale und die formal-pragmatische (Ryan, 2007, 29), (Ryan, 2012a, §17). Ryan folgend können die insgesamt acht konstitutiven Aspekte von Erzählungen wie ein Werkzeugkasten genutzt werden, um zur eigenen Begriffsdefinition zu kommen. Für einen engen Begriff von Erzählung würden die drei folgenden Eigenschaften ausreichen (Ryan, 2007, 30).

Spatial Dimension [Hervorhebungen im Original, Anm. MKS]

(1) Narrative must be about a world populated by individuated existents.
Temporal Dimension
(2) This world must be situated in time and undergo significant transformations.
(3) The transformations must be caused by non-habitual physical events. (Ryan, 2007, 29).

Bei diesem eigenschaftsbasierten Definitionsansatz steht also das Individuum in Raum und Zeit im Mittelpunkt, dessen (physische) Erfahrungswelt sich signifikant und auf ungewöhnliche Weise verändert. Dabei ist die Veränderbarkeit der dargestellten Welt eine besondere Qualität der Erzählung: Eine statische Welt wird

beschrieben, eine sich verändernde Welt wird erzählt (Ryan, 2007, 27). Zwar ist der Weltbegriff bei Ryan nicht allein auf räumliche Aspekte von Erzählungen beschränkt, die grundsätzliche Dynamik, die sie beschreibt, schließt aber auch die räumliche Dimension mit ein. An dieser Stelle bietet sich ein Anknüpfungspunkt an die Ausführungen zum literarischen Raum in Abschnitt 2.2.1. Wenn Erzählungen grundsätzlich eine veränderbare Welt zum Gegenstand haben, so muss auch der darin dargestellte Raum grundsätzlich dynamisch konstruiert werden. Er ist also kein festes Konstrukt, sondern ebenso veränderbar wie die gesamte storyworld.

Meister (2017) macht sieben fundamentale Funktionen des Erzählens aus: Darstellung, Handeln, Konstruktion, Exploration, Introspektion, Kognition und Reflexion. Alle diese Aspekte sind auch für die Darstellung von Raum bedeutsam, nur zwei können allerdings direkt auf eine intratextuelle Betrachtung angewendet werden, wie sie im Zentrum dieser Studie steht. Dass Orte und Räume in Erzählungen repräsentiert werden können, ist geradezu die Grundvoraussetzung. Dass sie aber auch Teil sozialer und historischer Identitäten sein können und dass diese sowohl auf der Ebene der Schreibenden und Lesenden als auch im Text selbst wirksam werden, ist vielleicht weniger offensichtlich. Meister bezieht diese Identitäten konstruierende Funktion zwar auf menschliche Individuen und Gruppen, die (Aspekte von) Erzählungen in ihre eigene Biographie oder die Gruppenidentität integrieren können (Meister, 2017, 101). Doch die Ausführungen in Abschnitt 2.3.3 haben gezeigt, dass soziale und kulturelle Faktoren auch für Orte und Räume von Bedeutung sind. Diese Dimension findet sich auch in Erzähltexten, in denen ein Haus z. B. nicht einfach nur ein Haus, sondern das Zuhause einer Figur sein kann – es kann ein Haus in einer angesehenen Gegend oder einem zwielichtigen Milieu sein usw. Der Kultursektor zeigt, dass solche textbasierten identifikatorischen Aspekte von Räumen sogar auf realweltliche Orte (zurück)wirken können, wie z. B. der Fall beim „Buddenbrook-Haus“ in Lübeck (vgl. dazu z. B. Selbmann, 2018; Lipinski, 2019). Die anderen von Meister aufgezählten Funktionen des Erzählens als Interaktion (Handeln) und als Möglichkeit zur Erkundung von Handlungsmöglichkeiten (Exploration) beziehen sich beide fundamental auf die Erzählenden, die Exploration auch auf Rezipierende. Die Teilhabe am fremden Wahrnehmen (Introspektion), das Erzählen als mentales Modell (Kognition) und die narrative-identity-These (Reflexion) – all diese Funktionen von Erzählungen können nicht ohne Einbezug eines rezipierenden Individuums analysiert werden. Dennoch können auf all diesen Funktionsebenen auch Orte und Räume betrachtet werden. Da das Forschungsdesign dieser korpusbasierten Untersuchung das aber nicht zulässt, muss eine solche Betrachtung hier ausgeschlossen und auf Anschlussstudien verschoben werden.

Die bisherigen Betrachtungen der Definitionsansätze von Erzählungen deuten bereits an, warum sich gerade Erzähltexte als Forschungsgegenstand dieser Studie

besonders eignen: Die dynamische, mehrdimensionale Konstruktion von Welten und damit auch von Raum ist ein genuines Potential von Erzählungen. Auf dieses Potential von Erzählungen verweist auch Meister (2017), indem er feststellt:

> Der Mehrwert des Erzählens gegenüber anderen Formen […] entspringt prinzipiell nicht der exakt verweisenden, sondern der relationalen, verknüpfenden Dimension der Aussagen, aus denen sie gebildet werden. (Meister, 2017, 98)

Da genau dieser Aspekt zentral für das entwickelte Raummodell ist, kommen andere Textformen wie Beschreibungen oder Argumentationen, die eine statische Welt darstellen, als Forschungsgegenstand nicht in Frage.

Der Fokus dieser Studie liegt auf räumlichen Aspekten von Erzählungen, die auf der Wortebene nachgewiesen werden können. Dabei ist eine fundamentale Grundannahme, dass Erzählungen semiotische Objekte sind, die auf elementaren Erfahrungen[1] aufbauen, ebenso wie jegliche Darstellung von Raum auf tatsächlicher Raumerfahrung basiert. Die Darstellung von Raum in literarischen Texten kann aber auch auf außertextliche Referenzwelten einwirken. Gegenstand der Untersuchung sind Romane, also eine schriftliche Form verbal vermittelter Erzählungen. Dass in dieser Analyse fiktionale Erzählungen im Fokus stehen, ist dadurch begründet, dass diese meist unabhängiger (vor allem im Hinblick auf Raum und Zeit) von konkreten Situationen der Produktion und Rezeption wirken müssen als faktuale Erzählungen. Eine erfahrungsbasierte Raumvorstellung kann also nur dann evoziert werden, wenn die Darstellung auf besondere Weise markiert wird (z. B. durch besonders häufige Nennung räumlicher Aspekte oder besonders ausführliche Beschreibungen der storyworld). Der Akt des Erzählens, also die Narration im Genette’schen Sinne, ist hier nicht weiter relevant. Damit werden bestimmte räumliche Aspekte von Erzählungen bewusst ausgeschlossen. Das sind vor allem der Raum des Autors als Schreibraum und der Raum des Lesenden als Leseraum. Das Zusammenspiel von histoire und récit bzw. story und discourse oder Inhalt und Form von Erzählungen ist besonders zentral, da betrachtet wird wie Rauminformationen referenziert werden und welche Räume im diachronen Verlauf von besonderer Bedeutung sind. Als Arbeitsdefinition von Erzählung gilt darum:

Definition 3.1 Eine Erzählung ist ein semiotisches Objekt, in dem ein besonderes Ereignis oder eine Reihe besonderer Ereignisse innerhalb einer storyworld dargestellt wird. Die storyworld wird im Verlauf der Erzählung dynamisch konstruiert und kann im Ganzen entweder physisch manifest oder mental sein oder einzelne

[1] Gemeint sind hier grundlegende Erfahrungen wie die Wahrnehmung von Raum und Zeit oder der Veränderbarkeit der Welt, nicht spezifische Einzelerfahrungen von Ereignissen.

Passagen von Erzählungen können als vorgestellt oder erträumt dargestellt werden. Die Geschichte einer Erzählung folgt einem bestimmten strukturierten Aufbau und verfügt über unterschiedliche, voneinander unterscheidbare Dimensionen (wie z. B. die räumliche oder zeitliche Dimension), die in einzelne Einheiten (wie z. B. Orte, Relationen, relationale Verben, Raumthemen, Raumbeschreibungen und Raumhinweise) unterteilbar sind. Erzählungen konstruieren soziale Identitäten, die bereits in diesen einzelnen Einheiten angelegt sein und auch auf eine außertextuelle Referenzwelt einwirken können.

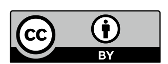
3.2 Definitorische Grundlagen narrativer Raumdarstellung

Als letzter definitorischer Schritt muss nun noch das entwickelte Raummodell mit der Arbeitsdefinition von Erzählung zusammengeführt werden. Die einzelnen Kategorien des Modells – Orte, Relationen, relationale Verben, Raumthemen, Raumbeschreibungen und Raumhinweise (und, wenn auch nur zu Anfang der Untersuchung, Raummetaphern) – zeigen sich auf unterschiedliche Weise in der Erzählung. Während Orte und Relationen meist mit Ausdrücken benannt werden, deren primäre Bedeutung räumlich ist, können relationale Verben, Raumbeschreibungen, Raumhinweise und Raumthemen mit Wörtern und Phrasen vermittelt werden, die neben räumlichen auch andere Aspekte der Erzählung tragen. Im ersten Abschnitt dieses Kapitels wurde bereits zusammengefasst, was eine Erzählung ist, auf der Arbeitsdefinition von Erzählungen aufbauend, gilt:

Definition 3.2 Die Darstellung von Raum in Erzählungen beinhaltet die explizite Nennung von Orten und Relationen, die primär Rauminformationen referenziert, und / oder von Ausdrücken, die neben ihrer primären Bedeutung auch Raumbeschreibungen, Raumhinweise, Raumthemen oder die Veränderung räumlicher Relationen implizieren. Sämtliche Kategorien der Raumdarstellung konstituieren dynamisch und auf Basis menschlicher Raumerfahrung zusammen den Raum der storyworld.

Die oben angeführte Arbeitsdefinition ist so angelegt, dass sie ausschließlich die inhaltlichen Aspekte literarischer Raumdarstellung (auf Ebene von story und discourse) abdeckt. Chatmans Begriffspaar des realen und des ästhetischen Objektes (Chatman, 1993, 26–27) folgend, findet also nur Letzteres Berücksichtigung. Alle physischen Objekteigenschaften – wie z. B. im Falle von Romanen der Umfang eines Buches, der Textraum oder der Schriftsatz (Mehr darüber z. B. in Lotman,

1973, 300–305; Ryan et al., 2016, 222–272 oder Mitchell, 1980, 544) – werden bewusst ausgeschlossen, um die Möglichkeit der Übertragung auf andere Arten von Erzählungen in Anschlussuntersuchungen zu gewährleisten. Das Raummodell deckt die Referenzierung von Rauminformationen ab, die mit spezifischen Verfahren narrativer Codierung auf der Wortebene bzw. auf Ebene von Textpassagen evoziert wird. Diese Verfahren müssen zunächst ausfindig gemacht werden, damit sie dann quantifiziert und analysiert werden können. In Abschnitt 2.2 wurde darauf hingewiesen, dass Raum in Erzähltexten eine Reihe narrativer Funktionen haben kann. Diese Funktionsebene ist im hier entwickelten Raummodell nicht enthalten. Dies ist vor allem dadurch begründet, dass die narrative Funktion von Raumdarstellungen stets einen Bezugspunkt benötigt (Nünning, 2001, 606), der selbst nicht unbedingt räumlich ist. Eine narrative Funktion von Raum kann z. B. sein, einer Erzählung einen Rahmen zu geben (Bal and van Boheemen-Saaf, 1985, 136–137). In diesem Falle wäre also Raum in Bezug zur gesamten Erzählung gesetzt. Raum kann auch eine strategische narrative Funktion haben (Ryan et al., 2016, 360–365). In diesem Falle kann Raum z. B. ermöglichen, dass eine Figur es leichter oder schwerer hat, an ihr Ziel zu gelangen, Raum steht dann in direktem Bezug zur Figur. Um die Variablen dieser Untersuchung möglichst homogen zu halten und nicht noch Parameter wie Figuren, Gesamterzählung oder sogar den Bezug zu einer außertextlichen Realität einzubeziehen, werden die narrativen Funktionen von Raum hier nicht weiter betrachtet. Eine Auswertung der Untersuchungsergebnisse im Hinblick auf die Funktionalität narrativen Raumes bietet sich allerdings für Anschlussuntersuchungen an.





4 Annotationsguidelines

Bis hierher wurde – abgesehen von ein paar wenigen konkreten Beispielen – abstrakt beschrieben, inwiefern Raumdarstellungen aus erzähltheoretischer Perspektive für Erzähltexte grundlegend sind. Um die Indikatoren narrativer Raumdarstellungen in Erzähltexten möglichst präzise erkennen und digital annotieren zu können, muss nun genau definiert werden, wie sich die unterschiedlichen Dimensionen des Raummodells im Text manifestieren können. Das Problem der metaphorisch verwendeten Raumausdrücke wird zwar im Modell selbst nicht abgebildet, muss aber berücksichtigt werden, um später bei der Messbarmachung von Raum in Erzähltexten wieder aufgegriffen (und gelöst) werden zu können. Für die sechs Kategorien von Indikatoren narrativen Raumes und für Raummetaphern werden darum in diesem Kapitel ausführliche Annotationsguidelines vorgestellt, die mithilfe der Best-Practice-Empfehlungen von Reiter (2020) erstellt wurden. Die Definitionen werden von literaturwissenschaftlichen Raumbegriffsbestimmungen gestützt oder von diesen abgegrenzt. Bei der Beschreibung der Kategorien wird jeweils eine Zugehörigkeitsbestimmung zu einem engen, weiten oder metaphorischen Raumbegriff vorgenommen. Das Kategoriensystem kann in Form eines Annotationstagsets nachgenutzt werden und wird im diese Arbeit ergänzenden GitHub-Repository als CATMA-Tagset OpenSource zur Verfügung gestellt[1].

[1] Im GitHub Repository zu finden unter dem Dateinamen „CATMA-Tag-Library_Export-2021-06-17T12_44_34.628.xml“ (Schumacher, 2021c).

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_4

Annotationseinheiten: Annotiert wird immer die kleinste mögliche Referenz auf ein Raum-Phänomen. Dies ist in der Regel ein Wort. In seltenen Fällen kann mehr als ein Wort annotiert werden. Besteht z. B. aufgrund der Position im Satz eine Trennung eines relationalen Verbs in zwei Worte, so werden beide als relationales Verb annotiert. Dasselbe gilt, wenn ein Ortsname aus zwei Wörtern besteht, wie z. B. „New York". Eine Ausnahme-Kategorie ist die der Raummetaphern. Diese bestehen vergleichsweise selten nur aus einem Wort (obwohl auch das möglich ist, z. B. wenn von der Größe einer Person gesprochen wird, damit aber nicht die Körpergröße, sondern eher die Güte gemeint ist). Meistens bestehen Metaphern, die Analogien zu Räumen nutzen, aus mehreren Wörtern und bilden kurze bis mittellange Phrasen. Nicht selten werden sie über einen ganzen Satz oder gar mehrere Sätze entwickelt. In diesem Fall werden alle zum Satzteil, Satz oder den Sätzen gehörenden Wörter als Teil der Raummetapher annotiert.

4.1 Orte

Grundsätzlich verstehe ich als Ort alles, was als fester, unbeweglicher Punkt oder als feste, unbewegliche Größe beliebiger Ausdehnung innerhalb der Topologie der storyworld identifiziert werden kann. Diese für diese Studie letztgültige Definition von Ort baut auf der im ersten Teil dieser Arbeit angenommenen Unterscheidung von Ort und Raum nach Descartes auf und verbindet diese mit literaturwissenschaftlichen insbesondere narratologischen Aspekten, die die Operationalisierung für Erzähltexte ermöglichen. Indikatoren dafür auf Wortebene sind Ortsnamen, (objektive, d. h. subjektunabhängige) Richtungsangaben und positionsangebende Adverbien. Auch in Unbestimmtheitsfällen können Orte impliziert werden. Leerstellen und nicht-Orte werden dagegen nicht zur Kategorie der Orte klassifiziert.

Das Aufrufen von Ortsnamen, wie Piatti es z. B. im Zusammenhang mit „ready-made-settings" erwähnt (Piatti, 2008, 232), ist die wohl expliziteste Erwähnung von Raum in Erzähltexten. Dennerlein hat detailliert herausgearbeitet, dass diese expliziten Ortsreferenzen über Eigennamen, Toponymika und Gattungsbezeichnungen funktionieren (Dennerlein, 2009, 77)[2]. Die philosophischen Betrachtungen in Abschnitt 2.3 haben gezeigt, dass Richtungen (*oben, unten, Himmel, Erde, Norden,*

[2] Dennerlein unterschiedet nicht zwischen Räumen und Orten, sondern zwischen Räumen und Ortsangaben. Was ich als Ort verstehe, ist bei ihr bereits ein Raum. Mit Ortsangaben bezeichnet sie dagegen Konkretisierungen wie z. B. „vor dem Haus" (Dennerlein, 2009, 209).

Osten, Westen und *Süden)* ebenfalls als konkrete Ortsreferenzen fungieren können, wenn sie objektiv, im Sinne von subjektunabhängig, gemeint sind. Aus Genettes Ausführungen zu „Wirklichkeitseffekten" (Genette und Knop, 1998, 118) lässt sich schlussfolgern, dass Raum zum Teil besonders detailliert beschrieben wird. Anhand seines Beispiels der „Ufer des vieltosenden Meeres" (Genette und Knop, 1998, 118) wird deutlich, dass diese detaillierte Beschreibung mit einer Häufung von Ortsnennungen einhergehen kann. „Ufer" und „Meeres" stehen hier als zwei Gattungsbezeichnungen nach Dennerlein (s. o.), die aber nur einen Ort (detailliert) beschreiben. Doch Orte können nicht nur explizit aufgerufen werden. Sie können auch die Form von Unbestimmtheitsstellen und Leerstellen (Piatti, 2008, 158–161) annehmen. Unbestimmtheitsstellen können durch die lesende Person gefüllt werden, z. B. wenn zu einem späteren Zeitpunkt in einer Erzählung der unbestimmte Raum konkretisiert wird (Piatti, 2008, 158–161). Auch Raumdarstellungen mit mehreren impliziten Hinweisen auf einen Ort können von Rezipierenden erfasst werden. Viehhauser (2020, 374) nennt als Beispiel den Satz „Ich erwachte und blickte der Ärztin ins Gesicht", in dem Lesende sofort das implizierte Krankenzimmer, evtl. auch ein Krankenhaus erkennen, das aber unbenannt bleibt. Unbestimmtheitsstellen können sich aber auf Ebene des Textes, z. B. in Beschreibungen oder Gattungsbezeichnungen wiederfinden, die lediglich später und nicht als erste Referenz auf einen Ort genannt werden (in Viehhausers Beispiel könnten Krankenzimmer und / oder Krankenhaus demnach in einem anderen Satz innerhalb der Erzählung direkt benannt werden), oder können über andere Kategorien literarischer Raumdarstellung erfasst werden (in Viehhausers Beispiel würden z. B. der Blick und das Gesicht der Ärztin als Raumhinweise erkannt und das Wort „ins" als Relation, es gibt also in dem Beispielsatz eine Reihe von Rauminformationen, auch wenn das raumthematisch wirkende „Krankenzimmer" unerwähnt bleibt). Auch in Fällen von Unbestimmtheiten können sich also Rauminformationen finden, die Teil des Raummodells sind. Leerstellen bleiben hingegen leer, können aber trotzdem von Bedeutung für Erzählungen sein, da sie z. B. Sprünge in der Handlung kennzeichnen können (Piatti, 2008, 158–161). Bei Leerstellen geht es aber gerade darum, dass sie nicht im Text auffindbar sind. Sie können also auch nicht in der mentalen (Re-)Konstruktion des konkreten Raumes der storyworld beim Lesen integriert werden. Sie bleiben aus diesem Grund im Modell literarischen Raumes unberücksichtigt. Das gleiche gilt für nicht-Orte wie z. B. „nirgends".

Definition 4.1 Orte sind konkret lokalisierbare Einheiten. Sie stehen fest und relativ unveränderbar an einer Position, die in einem geographischen Koordinatensystem lokalisiert werden könnte. Häuser, Zimmer, Berge, Landschaften u. Ä. werden als

Orte verstanden und annotiert. Als Orte werden i. d. R. ein bis drei Wörter annotiert, meistens handelt es sich um Nomen.

Beispiele:

- Er ging auf die *Bergspitze*
- Sie stand im *Zimmer*
- Sie wohnten in *New York City*

4.2 Räumliche Relationen

Indikatoren der Darstellung von räumlichen Relationen fallen in zwei Kategorien – Relationen und relationale Verben. Von besonderer Bedeutung sowohl für die Nennung von Orten als auch für die räumlichen Relationen sind die Äußerungen der Erzählinstanz und deren Perspektive. Schmid (2014, 121 ff) beschreibt in seinem Perspektivmodell Raum als einen von fünf Parametern:

> Die räumliche Perspektive wird konstituiert durch den Ort, von dem aus das Geschehen wahrgenommen wird, mit den Restriktionen des Gesichtsfelds, die sich aus diesem Standpunkt ergeben. Der Begriff der räumlichen Perspektive erfüllt als einziger der Termini, die einen Bezug des Erfassens und Darstellens auf ein Subjekt ausdrücken, die Intension von *Perspektive* oder *point of view* [Hervorhebungen im Original, Anm. MKS] im eigentlichen, ursprünglichen Sinne des Wortes. Alle anderen Verwendungen des Perspektivbegriffs sind mehr oder weniger metaphorisch. (Schmid, 2014, 123)

Ein Erzähler kann die gesamte storyworld kennen und den Lesenden in einer Art Draufsicht mit all seinen konkreten Orten und deren Relationen zueinander präsentieren, Ryan et al. sprechen (ähnlich wie de Certeau s. Abschnitt 2.3.3) vom „Map-View" und entwickeln damit Stanzels Begriff der „Perspektivierung" weiter (vgl. Ryan et al., 2016, 6 und 39). Oder die Erzählinstanz kann sich mit den Figuren und damit figurenähnlich durch den Raum bewegen. In diesem Falle kann sie das Gesamtgefüge entweder kennen und wie in einer Tour („Tour-View" nach Ryan et al., 2016, 7 und 27) durch den Raum präsentieren oder diesen ebenso lückenhaft und vielleicht sogar unstimmig wiedergeben wie Stanzel (Stanzel, 2008, 159–161) es für seinen Begriff der „Aperspektivierung" beschreibt. Konkret werden diese räumlichen Relationen, die häufig in der Erzählerrede enthalten sind, entweder durch Ausdrücke bezeichnet, die die Relation als solche bezeichnen oder durch Verben, die Relationen oder Veränderungen dieser implizieren können. Relationen können z. B. durch deiktische Ortsangaben bezeichnet werden. Dazu gehören Ausdrücke

wie „zwischen“, „links von“ usw. (Dennerlein, 2009, 209). Auch implizite Informationen über Relationen sind möglich, z. B. in Form von Präpositionalphrasen. Relationale Verben sind Verben der Bewegung wie „kommen“, „gehen“, „rennen“, aber auch Verben der Wahrnehmung wie z. B. „sehen“ oder „hören“, die nach Husserl Sicht- und Hörachsen bezeichnen, die Informationen über den Raum enthalten (vgl. Abschnitt 2.3).

4.2.1 Relationen

Definition 4.2 Als Relationen werden Beziehungen zwischen mindestens zwei Orten, Figuren oder Objekten verstanden. Mindestens einer bzw. eine oder eines dieser beiden Orte, Figuren oder Objekte muss Erwähnung finden, damit etwas als Relation annotiert wird. Annotiert wird allerdings nur das Wort, das für die Beziehung steht. Häufig werden Relationen in Form von Deiktika ausgedrückt. Auch Nomen, die für eine Relation per se stehen, wie z. B. „Distanz“, werden als Relation annotiert. Es handelt sich überwiegend um ein-Wort-Annotationen.

Beispiele:

- Sie stand *auf* dem Dach.
- Dieser Kirchturm war viel *höher*.
- *Zwischen* ihnen gab es viele Missverständnisse.

4.2.2 Relationale Verben

Definition 4.3 Als relationale Verben werden alle Bewegungsverben, Verben der Wahrnehmung und sonstige Verben, die ausdrücken, dass sich Relationen verändern, annotiert. Sind diese Verben mit Deiktika verbunden, wie z. B. bei „weggehen“, so werden im Falle einer satzbedingten Auseinanderschreibung (z. B. „sie ging weg“) beide zum Verb gehörigen Teile annotiert. Neben expliziten Richtungsverben werden auch alle Ausdrücke als relational interpretiert, die eine Relation implizieren. Dazu gehört z. B. „aussehen“, da der Blick einer Person hier auf eine andere fällt oder „rufen“, da eine Distanz überwunden wird. Relationale Verben werden nicht als solche annotiert, wenn sie zu einem Raumthema gehören. Nicht als relationales Verb annotiert wird z. B. „um nach ** *zurückzureisen*“, stattdessen wird „zurückzureisen“ hier als Raumthema annotiert.

Beispiele:

- Er *ging* in die Natur.
- Wir kamen überein, uns nicht voneinander zu *trennen*.

4.3 Hinweise auf den Raum

In der Literaturwissenschaft werden eine Reihe von Möglichkeiten beschrieben, um implizit auf den Raum hinzuweisen: Über die Wahrnehmung von Figuren, deren Positionierung, deren Bewegungen, die Erwähnung körperlicher Details und die Benennung von Objekten. Sowohl Erzähltheoretiker*innen (vgl. z. B. Hoffmann, 1978, 3–6; Bal and van Boheemen-Saaf, 1985, 136; Dennerlein, 2009, 150–160) als auch Tuan als Geograph weisen darauf hin, dass Raum nicht nur eine physikalische, sondern auch eine Wahrnehmungskategorie ist (Tuan, 1979, 34–35), was im Umkehrschluss bedeutet, dass die Raumwahrnehmung (sei sie nun optisch, auditiv oder taktil) von Figuren oder der Erzählinstanz Hinweise auf den Raum geben kann. Hört eine Figur oder eine Erzählinstanz in einem Roman Kirchenglocken, so bedeutet dies eine relative Nähe („Hörweite") zur Kirche. Diese Art von Rückschlüssen auf die Position einer Wahrnehmungsinstanz kann ungefähr sein, wie im vorangehenden Beispiel oder sehr konkret wie in einem von Dennerlein angeführten Beispiel, in dem die optische Perspektive einer Figur im stream of consciousness eine ziemlich genaue Positionierung zulässt (Dennerlein, 2009, 150). Nach Husserl (s. Abschnitt 2.3.2) ist es vor allem die Ebene der Sichtachsen, die Hinweise auf den Raum gibt. Erwähnungen von Augen und Blicken tragen also auch dazu bei, dass etwas über den Raum erzählt wird. Neben der Wahrnehmung (die Rückschlüsse auf die Position einer Figur zulässt) ist die explizite Positionierung von Figuren eine wichtige räumliche Komponente, da Figuren eine räumliche Körperlichkeit besitzen und sie ein wichtiger Bezugspunkt im System des relationalen Raumgefüges sind (Ryan et al., 2016, 6). Auch die Einstellungen, Handlungen, Aussagen (Nünning et al., 2004, 49) und Bewegungen (Bal and van Boheemen-Saaf, 1985, 140) von Figuren können Hinweise auf den Raum geben. Sowohl den Ausführungen Husserls zur Verortung von Gefühlen in einem Ich-Ding oder Ichleib (Husserl, 1991, 162), als auch der Idee vom Raum als res extensa (Descartes, 2007, 91,93 und 95) und letztendlich auch der Betonung des menschlichen Körpers als Wahrnehmungsinstanz von Raum (s. o.) folgend, wird die Darstellung körperlicher Aspekte von Figuren auch darauf geprüft, ob sie Hinweise auf Raum enthält. Das bedeutet, dass das gesamte Wortumfeld von Figurennennungen, insbesondere Darstellungen sinnlicher Wahrnehmung, für die literarische Darstellung von Raum von Bedeu-

tung sein kann. Butor stellt fest, dass auch die Nennung bestimmter Möbel oder Gegenstände Hinweise auf den Raum geben können. Er nennt diese Art räumlicher Kategorisierung „Dekor“ (Butor, 1968, 83–84).

Definition 4.4 Als Raumhinweise werden Ausdrücke annotiert, die zwar Informationen zum Raum geben, die aber keine eindeutige Aussage darüber machen, welcher geographische Raum oder welche räumliche Ausdehnung genau gemeint ist. Als Raumhinweise werden zum Beispiel Möbelstücke annotiert. Maßeinheiten wie Meter, Meilen u. Ä. geben ebenfalls Raumhinweise, da sie zwar eine genaue Länge angeben, nicht aber zwingend enthalten, wo sich diese (geographisch) erstreckt. Auch physische Merkmale wie Körperteile werden als Raumhinweise interpretiert. Der Kopf sitzt z. B. immer relativ weit oben, ist aber natürlich beweglich, die Hand ist häufig ein Raum-überbrückendes Element (z. B. beim Hände Reichen oder Greifen von Dingen) usw. Den in Abschnitt 2.3.2 beschriebenen Betrachtungen Husserls folgend, werden als Raumhinweise auch vage Verortungen im Innern von Figuren annotiert. Ebenfalls zurückgreifend auf die *Dingvorlesung* Husserls können auch Blicke, Geräusche und Ertastetes Hinweise auf den Raum geben. Darüber hinaus werden auch Ausdrücke als Raumhinweise gewertet, die nur vage Rauminformationen enthalten, wie z. B. die Wendung „in seinem *Blickfeld*“. Die Kategorie „Raumhinweis“ ist sehr weit gefasst und gehört zu einem sehr weiten Raumbegriff.

Beispiele:

- Sie verloren sich im *Gedränge*.
- Sein *Blick* fiel auf den Kirchturm.
- Sie waren viele *Meilen* gegangen.

4.4 Raumbeschreibungen

Zum „Dekor“ nach Butor gehören aber nicht nur Objekte, sondern auch deren adjektivische Beschreibungen (Butor, 1968, 83–84). Solche Raumbeschreibungen können nicht für sich allein stehen, sondern beschreiben immer ein Objekt in seiner Räumlichkeit. Raumbeschreibungen wie z. B. „glatt“ oder „klein“ weisen relativ explizit auf Raumdarstellungen hin. Raumbeschreibungen können aber auch zu den impliziten Indikatoren narrativer Raumdarstellung gehören. Das Adjektiv „stark“ wie in „mit starkem Arm“ weist z. B. implizit darauf hin, dass der beschriebene Arm recht muskulös ist und darum im Vergleich zu anderen Armen mehr Raum einnimmt (vgl. Chatman, 1993, 102 und Abschnitt 2.2.3).

Definition 4.5 Als Raumbeschreibung werden Adjektive annotiert, die entweder Räume beschreiben oder in einer Beschreibung eine räumliche Verknüpfung implizieren. Nicht als Raumbeschreibung annotiert werden beschreibende Ausdrücke, die Verbformen, wie z. B. das Gerundium nutzen. Diese werden als relationale Verben annotiert, wie z. B. bei „das *geschwungene* Dach". Raumbeschreibungen können mehr oder weniger explizit ausgedrückt werden, sodass die Kategorie mehr Ausdrücke umfasst als ein enger Raumbegriff einschließen würde. Die Kategorie Raumbeschreibung gehört darum zu einem erweiterten (aber nicht unbedingt sehr weiten) Raumbegriff.

Beispiele:

- Es war ein *großes* Zimmer.
- Das *italienische* Kraut.
- Sie war von *schmaler* Gestalt.

4.5 Raumthemen

Die kulturelle Codierung von Räumen kann zur metaphorischen Nutzung von Raumbegriffen führen. Darum wurden literaturwissenschaftliche Ansätze, die diesen Aspekt mit einbeziehen vielfach als metaphorische Raumdeutung bezeichnet (s. Abschnitt 2.2.1). Allerdings unterscheiden z. B. Ryan et al. zwischen einer rein metaphorischen Nutzung von Raumbegriffen und einer semantischen Verwendung, bei der der Bezug zu konkreten Räumen erhalten bleibe (Ryan et al., 2016, 17). Durch literarische Konventionen, die z. B. an ein Genre geknüpft sein können, werden bestimmte Raumausdrücke zu thematischen Symbolen. So stehe im Märchen z. B. das Schloss für Macht, in Sagen der Berg für das Heim des Göttlichen (Ryan et al., 2016, 37). Bal nennt diese literarischen Raumsymbole „Topoi" (Bal and van Boheemen-Saaf, 1985, 141)[3]. Die Symbolik wird interpretativ bzw. in den Beispielfällen durch Genrewissen erkannt. Nach Lotman gibt es Topoi, die für narrative Texte von besonderer Bedeutung sind, dazu gehört die Grenze (Lotman, 1973, 328–329). Für die Beschreibung von Grenzen sind Nennungen bestimmter Objekte von besonderer Bedeutung, wie z. B. Fenster, Türen und Treppen (Hoffmann, 1978, 470–474). Fahrzeuge sind weitere Raumsymbole, da sie das literarische Reisemotiv

[3] Grundsätzlich wird die rezeptionsästhetische Perspektive in dieser Untersuchung zwar nicht mit einbezogen, da Raumthemen aber sehr stark konventionalisierte semantische Felder umfassen, die in literarischen Texten immer wieder aufgerufen werden, ist lediglich ein sehr schematisches Wissen zur Decodierung notwendig.

unterstützen (Hoffmann, 1978, 470–474). Vom 18. Jahrhundert an gewinnt die Stadt als literarisches Raumsymbol an Bedeutung (Hoffmann, 1978, 388). Auch Erinnerungsräume (zu Erinnerungsräumen in der Literatur vgl. Bal and van Boheemen-Saaf, 1985, 140) wie z. B. Friedhöfe, die mit Trauer assoziiert werden, können als Raumsymbole in Erzähltexten fungieren. Hier überschneidet sich das Konzept des Erinnerungsraum mit dem der Heterotopien von Foucault (s. Abschnitt 2.2.2). Heterotope Räume werden ebenfalls als den Raumthemen zugehörig interpretiert. Sowohl bei Erinnerungsräumen als auch bei Heterotopien entsteht die semantische Ebene nicht allein durch literarische Konventionen, sondern durch eine nicht oder höchstens zum Teil literarische sozio-kulturelle Prägung. Der Begriff des Raumthemas soll hier also auch für semantisch aufgeladene Räume, die ihre Bedeutung im außerliterarischen Zusammenhang bekommen, verwendet werden. Raumthemen werden in literarischen Texten immer wieder aufgerufen und sind dadurch so stark konventionalisiert, dass sie intersubjektiv zugänglich sind und ohne intensive Interpretationsleistung wiedererkannt werden können (wird z. B. erwähnt, dass eine Figur sich im Zug befindet, so wird geradezu reflexartig das Raumthema der Reise aufgerufen, wird ein Sarg genannt, so steht dies in engem Zusammenhang mit dem heterotopen Raum des Friedhofs). Werden sie vom Rezipierenden nicht als Raumthemen erkannt, so können sie trotzdem innerhalb der storyworld eine raumthematische Funktion einnehmen (wie z. B. der Fall bei Wörtern wie „Tür", „Fenster", „Wand", die zum semantischen Feld des Raumthemas der Grenze gehören und für ein System der Öffnung und Schließung stehen). Das Aufrufen komplexer Topoi kann über simple Ortsnennungen funktionieren; die Erwähnung einer Nerven- und Heilanstalt ruft z. B. die Assoziation eines geschlossenen Raumes mit festgelegten Verhaltensmustern auf (Dennerlein, 2009, 178–189). Nünning und Nünning weisen darauf hin, dass diese semantischen Zuschreibungen vor allem, wenn sie sich auf vage Räume beziehen (z. B. Natur oder Heimat), keinesfalls immer gleichen Inhalts sein müssen (Nünning et al., 2004, 50–51). Hofmann schlägt eine Skala vor, die von geschlossener Raumsymbolik bis zur offenen Raumsymbolik reicht (Hoffmann, 1978, 319). Er benennt zehn Kategorien von Raumsymbolen, die unterschiedlich offen bzw. geschlossen sind (Hoffmann, 1978, 320–327). Wie in Abschnitt 2.2.1 bereits ausgeführt, hat Hofmann auch betont, dass Raumsymbolik dynamisch ist und sich innerhalb eines Textes wandeln kann. Das bedeutet auch, dass sie vom intratextuellen Kontext abhängig ist oder zumindest sein kann. Neben Topoi und Raumsymbolen können auch raumzeitliche Symbole für Raumthemen von Bedeutung sein. Dazu gehört z. B. die Uhr (Hoffmann, 1978, 368). Literaturgeschichtlich gibt es außerdem bestimmte prototypische Raumstrukturen, die mit konkreten Themen verknüpft sind, wie z. B. der Inselraum (Dennerlein, 2009, 189–192).

Definition 4.6 In die Kategorie *Raumthema* fallen Ausdrücke, die zu einem in der Literaturwissenschaft ausgemachten, vom Raum geprägten, Motiv beitragen. Unter dem Begriff Raumthema sind räumliche Motive wie das von Lotman als besonders wichtig hervorgehobene Grenzmotiv, zu dem auch Elemente der Öffnung und Schließung wie Türen, Fenster und Treppen gehören, Bachtins Chronotopoi wie der Reisetopos und Foucaults Heterotopien wie Begräbnisstätten, Feste und Schiffe zusammengefasst. Einzelräume wie Heterotopien gehen zum Teil in größeren Raumthemen auf. So werden z. B. nicht nur die als heterotop klassifizierten Schiffe als Raumthema annotiert, sondern alle Fahrzeuge und Fortbewegungsmittel. Annotiert wird in der Regel ein Wort, das den Bezug zum Raumthema herstellt. Die Raumthemen sind manchmal an konkrete räumliche Elemente gebunden, zum Teil aber auch abstrakter, wie z. B. das Thema der Freiheit. Da in jedem Fall die räumliche Thematik auf einen anderen Bedeutungszusammenhang übertragen wird, der z. B. eher emotional oder auch zeithistorisch bedeutsam ist, gehören Raumthemen zu einem weiten Raumbegriff. Raumthemen haben zum Teil eine metaphorische Bedeutungsebene (z. B. kann die äußere, physische Grenze für eine innere Begrenztheit stehen), sie sind aber meistens an physische Elemente geknüpft. In jedem Fall bietet sich eine Unterkategorisierung in einzelne Raumthemen für einen weiteren Schritt im Analyseprozess an.

Beispiele:

- Sie kam an eine *Tür*, zögerte aber, sie zu *öffnen*.
- Er betrat das *Schiff* und *segelte* davon.
- Sie mussten sie allzu früh *begraben*.

4.6 Annotation von Raummetaphern

Bisher wurde die Raummetaphorik nur am Rande erwähnt, da Ausdrücke ohne Bezug zum konkreten Raum nicht im Zentrum der Betrachtungen stehen. Um einschätzen zu können, inwiefern Metaphern ein Störfaktor der angewendeten digitalen Methodik sind (vgl. Schumacher, 2022) und wie sie in der Analyse berücksichtigt werden müssen bzw. können, ist es aber wichtig, an dieser Stelle festzuhalten, wie Metaphern hier verstanden werden. Die Besonderheit von Metaphern liegt darin, dass sie eine Referenzbeziehung herstellen, die nicht allein zwischen Wort und Objekt besteht, sondern eine weitere Größe einbezieht. Diese dritte Größe ist durch Ähnlichkeit mit den beiden ersten verbunden. Es entsteht also eine Dreiecksbeziehung zwischen Wort, Objekt und Ähnlichem (Wenz, 1997, 32). Nun können

Metaphern sowohl als rein sprachliche Stilmittel betrachtet werden oder als prägende Ausdrücke menschlichen Handelns (vgl. Wenz, 1997, 33 und Blumenberg, 1971, 213). Unter der letzteren Prämisse werden Metaphern nicht als einzelne Ausdrücke betrachtet. Stattdessen bündeln Vertreter der kognitiven Metapherntheorie wie Lakoff und Johnson mehrere Ausdrücke zu Metapherngruppen oder -systemen, die alle auf eine Kernmetapher zurückgeführt werden können (Lakoff, 1998, 9). Diese Systeme zeigen nicht nur kulturelle Bedeutungen (Lakoff, 1998, 15) auf, sondern wirken auch auf das Alltagsleben, Wahrnehmen und Handeln von Menschen ein (Wenz, 1997, 33). Blumenberg entwickelte eine wissenschaftliche Analysemethode, die sich hauptsächlich auf Metaphern konzentriert, die Metaphorologie (Blumenberg et al., 2013, 16–17). Auch Blumenberg geht davon aus, dass die Nutzung bestimmter Metaphern im zeithistorischen Kontext, auf Denksysteme, Wahrnehmung und Verstehen der Menschen verweist (Blumenberg et al., 2013, 14). Andersherum wirken sich stark genutzte Metaphern auf die Denksysteme und schließlich auch die Lebenswelt der Menschen aus; es gibt also eine Wechselwirkung zwischen Metaphorik, die genutzt wird, um Denksysteme zu beschreiben, und der Nutzung von Metaphern und diesen Denksystemen selbst (Blumenberg und Haverkamp, 2001, 195).

Dass die Wahrnehmung des Raumes eine essentielle menschliche Erfahrung ist, spiegelt sich auch darin wider, dass es eine große Anzahl von Raummetaphern gibt (Wenz, 1997, 33). In gegenständlichen Kontexten – wie auch der Untersuchungsbereich dieser Analyse einer ist – sind Metaphern allerdings ein Störfall (Blumenberg und Haverkamp, 2001, 194), da sie für das „Uneigentliche“ stehen (Blumenberg et al., 2013, 14), also für etwas, das in der Sprache nicht direkt gesagt wird.

In Abschnitt 2.1 wurde bereits festgehalten, dass räumliche Metaphern, die keinerlei Referenz auf Rauminformationen beinhalten, im entwickelten Modell keine Rolle spielen. Problematisch daran ist, dass räumliche Metaphern, die z. B. Ortsnamen einschließen, nicht von einer automatischen Erkennung ausgeschlossen werden können. Würde es z. B. in einem Erzähltext heißen, jemand trage Eulen nach Athen, so wäre es für eine Software nicht möglich zu erkennen, dass der Raumausdruck „Athen“ hier nicht wörtlich gemeint ist[4]. Das Wort selbst, die grammatische Kon-

[4] Das Beispiel „Eulen nach Athen tragen“ ist zwar eingängig, zeigt aber nur eine Form der stark konventionalisierten Raummetapher. Diese Art von Metaphern könnten mit vergleichsweise einfachen computergestützten Verfahren wie einem Listenabgleich erfasst werden. Leider gibt es derzeit aber noch kein digital vorliegendes Verzeichnis literarischer Raummetaphern, mit dem ein solcher Abgleich ad hoc durchgeführt werden könnte. Problematisch ist auch, dass diese Form der stark konventionalisierten metaphorischen Phrase in literarischen Texten vergleichsweise selten ist (wie sich während der manuellen Annotation des Trainingskorpus für das Machine Learning gezeigt hat). Häufiger sind Metaphern, die nur ein Wort umfassen,

struktion des Satzes, vorangehende und nachgestellte Wörter – all das trägt dazu bei, dass trainierte Algorithmen „Athen" hier eindeutig als Ort erkennen würden. Nur durch die Konventionalisierung dieser Raummetapher wissen Lesende hingegen sofort, dass der fiktive Charakter sich nicht mit Eulen beladen auf den Weg nach Athen macht. Diese Uneigentlichkeit der Metaphern ist es, die für diese Studie am problematischsten ist. Da bisher keine Verfahren zur automatischen Erkennung und damit auch kein automatischer Ausschluss von räumlichen Metaphern in deutschsprachigen Erzähltexten entwickelt wurde, konnten Metaphern als Störfaktor bei der Analyse nicht komplett ausgeschlossen werden. Die Tests des Raumerkennungstools haben aber gezeigt, dass dieser Störfaktor relativ gering ausfällt (vgl. Kapitel 6 und Schumacher, 2022). Um diese Bewertung vorzunehmen, mussten Metaphern im Machine-Learning-Training zunächst als eigene Kategorie mit einbezogen werden. Um Metaphern im Trainingskorpus als solche zu markieren, wurden sie in der Umfeldanalyse nach der MIPVU-Erkennungsmethode (vgl. Steen, 2010) annotiert. Statt sie aber wie die anderen Kategorien im endgültigen Trainingsprozess des Machine-Learning-Tools einzubeziehen, wurde für die im Trainingskorpus enthaltenen Metaphern ein paralleler Workflow entwickelt. Alle im Trainingskorpus enthaltenen Raummetaphern wurden mit Hilfe der Software Neo4J (Graphdatabase, 2022) in eine relationale Graphdatenbank eingepflegt. Die Metaphern wurden nach Umfang in die drei Kategorien „ein-Wort-Metapher", „metaphorische Phrase" und „komplexes Raumbild" unterteilt. Für jede Metapher wurde ein Kernwort identifiziert. Jede Metapher wurde mit dem Roman / den Romanen, in dem / denen sie vorkommt, und dessen / deren Entstehungszeit verknüpft. Die so entstandene Datenbank literarischer Raummetaphern (*laRa*)[5], die rund 800 Raummetaphern aus 80 Romanen aus vier Jahrhunderten umfasst, kann z. B. nach Metaphern, Entstehungszeit und Kernworten durchsucht werden. Mit Hilfe dieser Datenbank konnten bei

---

dass mal wörtlich und mal metaphorisch genutzt werden kann wie bspw. „Himmel". Oft werden auch neue räumliche Metaphern geprägt oder komplexe metaphorische Raumbilder erzeugt, die über mehrere Sätze aufgebaut werden – beides bleibt oft textspezifisch. Aus diesen Gründen wurde die Implementierung eines Listenabgleichs zunächst verworfen und stattdessen die Arbeit an einer Datenbank literarischer Raummetaphern, „laRa", als Index für Raummetaphern in literarischen Texten begonnen. Aus diesem Index kann perspektivisch ein Lexikon literarischer Raummetaphern generiert werden, sobald eine kritische Menge an Beispielen erreicht ist. Mit einem solchen digitalen Lexikon könnte dann ein Listenabgleich bei der Implementierung einer neuen Version des Raum-Classifiers oder gar eines neuen Classifiers, der spezifisch für literarische Raummetaphern ist, mit einbezogen werden.

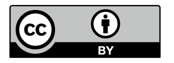
[5] Die Datenbank literarischer Raummetaphern laRa steht zur freien Nachnutzung zur Verfügung und kann als Neo4J-Dump oder in Form von Gephi-Tabellendaten über Zenodo herunter geladen werden (vgl. Schumacher, 2021b). In dem Repository zur Datenbank befinden sich neben den Daten auch die Dokumentation der verwendeten Texte.

der Analyse der Kernkorpora häufige Raummetaphern berücksichtigt werden, die sich als Störfaktor im Machine-Learning-Modell eingeschlichen haben.

Definition 4.7 Als Raummetapher werden alle Ausdrücke annotiert, die auf Wortebene Räume referenzieren, auf der Bedeutungsebene aber auf etwas anderes abzielen. Autoren von Erzähltexten können Raummetaphern neu einführen und tun dies häufig in vergleichsweise langen Passagen, in denen räumliche Bilder beschrieben werden. Außerdem gibt es Ausdrücke und Phrasen, die hochfrequent raummetaphorisch genutzt werden, wie z. B. „groß“ im Sinne von bedeutend oder die Phrase „etwas (selbst) in der Hand haben“. Schließlich gibt es noch Ausdrücke, die ursprünglich räumlich sind, aber nicht immer unbedingt einen physischen Raum bezeichnen müssen und so hochfrequent für nicht-Räumliches verwendet werden, dass der Gebrauch nicht mehr als metaphorisch bezeichnet werden kann. Dazu gehören z. B. Phrasen wie „das ist doch keine große Kunst“, die nicht als raummetaphorisch gekennzeichnet werden. Die Hauptschwierigkeit bei der Annotation räumlicher Metaphern besteht darin, diese Dynamik von Einführung, Etablierung und Übergang in den normalen Sprachgebrauch richtig zu erfassen, wie auch Thaller (2021) beschreibt. Als Raummetaphern können einzelne Wörter, Phrasen oder längere Passagen von bis zu mehreren Sätzen annotiert werden. Nicht annotiert wird z. B. „Sie hatte große Angst“, da mit „groß“ hier eine quantitative Beschreibung zu dem Wort „Angst“ hinzugefügt wird, die im Sprachgebrauch so etabliert ist, dass sie nicht mehr als metaphorisch kategorisiert werden kann.

Beispiele:

- Es ist die höchste Zeit, dass ich wieder auf mich selbst und *auf den Daumen-, Zeige- und Mittelfinger* des Werks zurückkehre.
- In einer wahren Geschichte das Wort *aus dem Munde zu nehmen*.
- Sie war eine *große* Persönlichkeit.


cc
BY

Teil II
Die Methode



5 Automatische Annotation

Die zentrale digitale Methode, die in dieser Arbeit angewendet wird, ist Named Entity Recognition (NER). Mit Hilfe dieses computerlinguistischen Verfahrens werden in der Regel klar benennbare Einheiten, wie z. B. Eigennamen, in Texten erkannt und annotiert (Schumacher, 2018a, §1). Die am häufigsten in NER-Tools implementierten Annotationskategorien sind Personen, Orte und Organisationen. Zum Einsatz kommt die frei verwendbare Software Stanford Named Entity Recognizer (StanfordNER) (Finkel et al., 2005), die auf einer Implementierung von Conditional Random Field (CRF) Algorithmen (vgl. Sutton and McCallum, 2010) beruht (Manning et al., 2014). Dabei handelt es sich um vektorbasierte Vorhersagealgorithmen, die auf Basis eines Machine-Learning-Prozesses aufgrund bestimmter Kontextinformationen im Text Annotationskategorien erlernen können (mehr dazu in Abschnitt 5.1). Diese Kernmethode wurde, wenn nötig, durch andere digitale Methoden ergänzt. Wie bereits erwähnt, wurde eine relationale Graphdatenbank in Neo4J angelegt, um Raummetaphern in die Analyse einbeziehen zu können. Um die Ergebnisse der automatischen Raumannotation auswerten und visualisieren zu können, wurde die Annotationssoftware CATMA[1] (Meister et al., 2018 und Gius et al., 2022) genutzt. Die Abfrageergebnisse dieses Tools wurden wiederum überführt in die Netzwerkanalysesoftware Gephi (Bastian et al., 2022), um zu untersuchen, inwiefern die Raumkategorien zusammen wirken. Auch das Topic-Modeling-Tool DARIAH TopicsExplorer (Simmler et al., 2019) wurde eingesetzt, um die Analyse der Annotationsdaten zu ergänzen. Insgesamt liegt der Studie also ein Mixed-Methods-Ansatz zu Grunde, der die Methoden der digitalen Annotation (vgl. Jacke, 2018), der Named Entity Recognition (vgl. Schumacher, 2018a), der Netzwerkana-

[1] Aufgrund leicht unterschiedlicher Funktionalitäten wurde sowohl CATMA 5 (Meister et al., 2018) als auch CATMA 6 (Gius et al., 2022) eingesetzt.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_5

lyse (vgl. Schumacher, 2018b), des Topic Modelings (vgl. Horstmann, 2018) und der Metadatenauswertung mit Hilfe relationaler Datenbanken umfasst.

Das Machine-Learning-Verfahren NER umfasst eine Reihe von Arbeitsschritten, für die unterschiedliche Parameter von Bedeutung sind, die in diesem und im folgenden Kapitel 6 genauer betrachtet werden. Im Fokus stehen zuerst die Grundlagen der Methode und die in dieser Arbeit eingesetzte Variante der NER mittels Conditional Random Fields Algorithmen (s.o.). Um die Idee der auf Wort- und Kontexteigenschaften beruhenden Erkennung bestimmter Kategorien umzusetzen, können diverse Features definiert werden, die beim Machine Learning berücksichtigt werden. Dabei handelt es sich um Worteigenschaften der Wörter, die in eine Annotationskategorie fallen, wie die Kombination bestimmter Buchstaben (oder auch Prä- und Suffixe) oder häufig vor oder nach einem Objekt einer Kategorie stehende Wörter. Die Bedeutung der Features für NER und die Wahl der in dieser Studie eingesetzten Feature-Kombination wird in Abschnitt 5.2 beschrieben und begründet. Die grundsätzliche Beschaffenheit der implementierten Algorithmen und die definierten Features bilden zusammen das Machine-Learning-Tool, also die Basis des Lernprozesses des Computers. Im nächsten Schritt kann dieser Lernprozess, das sogenannte Training, begonnen werden. Dazu müssen Trainingsdaten erstellt werden, eine kritische Masse von Beispielannotationen, aus denen mit Hilfe der CRF-Algorithmen und anhand der Features die Annotationskategorien erlernt werden. Ergebnis des Lernprozesses ist ein CRF-Classifier. In diesem Classifier sind die Regelmäßigkeiten, die bei der Auswertung der Beispielannotationen erkannt wurden, festgelegt. Zur Umsetzung der automatischen Erkennung der Wörter, die in eine der NER-Annotationskategorien fallen, sind sowohl das Tool als Basis der Erkennung als auch ein Classifier für die spezifischen Annotationskategorien notwendig. Nach Abschluss des Trainings folgt ein Schritt, bei dem die Ergebnisse der automatischen Annotation getestet werden. Wieder werden manuell Beispielannotationen erstellt, die dann mittels Tool und Classifier mit der automatischen Annotation abgeglichen werden. Zur Bewertung werden drei Kennzahlen errechnet: Precision, Recall und F1-Score. Mit Precision wird berechnet, wieviel Prozent der annotierten Wörter korrekt annotiert wurden (vgl. Powers, 2020, 38). Recall setzt den Anteil der annotierten Wörter ins Verhältnis zur Anzahl aller Wörter im Text, die in die Annotationskategorie(n) fallen (vgl. Powers, 2020, 38). Precision gibt also einen Eindruck davon, wie viel dessen, was annotiert wurde, korrekt ist (Genauigkeit), während Recall zeigt, wie viel von dem, was hätte annotiert werden sollen, tatsächlich annotiert wurde (Vollständigkeit). Der F1-Score ist ein mathematischer Wert, der die beiden Kennzahlen in angemessener Weise zusammenfasst (Mikheev et al., 1999, 2). Wie diese Tests bei derzeit als State of the Art geltenden

Tools ausfallen und inwiefern eine Domänenadaption für die Literaturwissenschaft bereits ähnliche Resultate erbringen konnte, wird in Abschnitt 5.2 herausgearbeitet.

Mit Hilfe von StanfordNER wurden bereits linguistische Modelle für die deutsche Sprache trainiert, die die zentralen NER-Kategorien Personen, Orte und Organisationen umfassen. Bis heute können diese Modelle als State of the Art bezeichnet werden, da sie mit 80 % F1-Score sehr gute Ergebnisse in der Erkennung erreichen (vgl. Faruqui and Padó, 2010; Hänig et al., 2014, 113). Darüber hinaus bietet StanfordNER eine einfach bedienbare graphische Benutzeroberfläche, was die Nachnutzung des hier entwickelten Raum-Classifiers erleichtert. Conditional Random Fields sind zwar nicht die einzigen Algorithmen, die in NER Tools zum Einsatz kommen können, die Wahl fiel hier aber dennoch auf eine Software mit CRF-Classifiern, da diese durch ihre Kontextsensitivität besonders gut auch auf Gegenstände angewendet werden können, die über die klassischen Named Entities hinausgehen.[2] Die Vor- und Nachteile dieser Wahl werden in Form einer Ausführlichen Methodenkritik in Abschnitt 5.3 diskutiert.

5.1 Grundlagen der Named Entity Recognition

Named Entity Recognition gehört zum Methodenkomplex des Natural Language Processings (NLP) und wird häufig als wichtige Vorbereitung anderer Aufgaben beschrieben (vgl. dazu z. B. Mikheev et al., 1999, 1; Faruqui and Padó, 2010, 1; Marrero et al., 2013, 482; Weber and Pötzl, 2014, 158). Dazu gehört z. B. die automatische Informationsextraktion oder Übersetzung (vgl. Faruqui and Padó, 2010, 1; Marrero et al., 2013, 482;). Obwohl er bereits 1996 zum ersten Mal verwendet wurde, ist der Begriff der Named Entity nach wie vor unterspezifiziert (Marrero et al., 2013, 482). Erklärungsansätze reichen von der sehr unspezifischen Aufzählung von Beispielen (Weber and Pötzl, 2014, 158) bis hin zu einer engen Definition als Einheit aus einem Objekt und dessen (möglichst) einzigartigem Bezeichner (Marrero et al., 2013, 484). Vielen in der Computerlinguistik entwickelten und eingesetzten NER-

[2] In neueren Studien zur Methode der Named Entity Recognition werden CRF mit Deep Learning, genauer mit Word2Vec-Implementierungen, verbunden. Die Ergebnisse sind für die Figurenerkennung in literarischen Texten mit Hilfe von NER sehr vielversprechend (vgl. Jannidis et al., 2015; Wiedemann et al., 2018; Li et al., 2020). Die bisherigen Tests machen allerdings noch nicht deutlich, ob eine Übertragbarkeit auf andere Kategorien ebenso gut durchgeführt werden kann wie mit reinen CRF-Classifiern. Anders ausgedrückt: Es wird in den bisherigen Studien nicht klar, ob die Kombination mit Word2Vec die Dehnbarkeit der Methode erhält. Da hier NER nicht mehr im eigentlichen Sinne verwendet wird, wird auf den Einsatz einer Word2Vec-Implementierung verzichtet.

Classifiern scheint eine solche, enge Definition von Named Entities inhärent zu sein, da meist nur Referenzen berücksichtigt werden, die mit Hilfe eines Eigennamens erzeugt werden (Kuhn, 2018, 26). Dieser Arbeit liegt die weitgehend anerkannte, aber über den Fokus auf Eigennamen hinausgehende, Beschreibung von Named Entity Recognition als Mittel, um sowohl bekannte als auch unbekannte Objekte bekannter Hierarchien in unbekannten Texten ausfindig zu machen, zu Grunde (Marrero et al., 2013, 484). Außerdem wird hier berücksichtigt, dass das Einsatzgebiet der Methode immer auch die Kategorienfindung mitbestimmt und auf die jeweilige Domäne angepasst werden sollte (vgl. Marrero et al., 2013, 487; Kuhn, 2018, 12 und 21). Obwohl beim Vergleich gängiger Tools auffällt, dass Personen, Orte und Organisationen die typischsten Kategorien der Named Entity Recognition sind (Mikheev et al., 1999, 1; Marrero et al., 2013, 484–485), gibt es auch solche, die sich auf eine oder wenige Kategorien beziehen (z. B. das NER-Modell für literarische Figuren von Jannidis et al., 2015) und solche, die die Methode dahingehend ausreizen, dass sie möglichst viele Kategorien automatisch annotieren lassen. Marrero et al. (2013, 485) berichten von Tools mit 39 und 45 Kategorien. Auch stark domänenspezifische NER-Systeme, z. B. zur Erkennung medizinischer Fachtermini für Krankheiten, wurden entwickelt und eingesetzt (Jiang et al., 2016, 21; Kuhn, 2018, 33). Die Diskussion über definitorische Grundlagen der NER und die unterschiedlichen Modelle, die bereits implementiert wurden, lässt den Schluss zu, dass die Methode sehr anpassbar ist und darum gut auf andere Zusammenhänge übertragen werden kann.

Grundsätzlich kann die Aufgabe von Named-Entity-Recognition-Software als zweischrittig verstanden werden. Zunächst muss das Tool grundsätzlich diejenigen Objekte identifizieren, die der vordefinierten Oberkategorie entsprechen (z. B. Named Entities) (Sutton and McCallum, 2007, 5). Dann müssen all diese Funde nach Art klassifiziert werden (z. B. Personen, Orte, Organisationen (Sutton and McCallum, 2007, 5). Die Erkennung erlernt das Tool in einem Machine-Learning-Training, bei dem die vordefinierten Features mit dem Trainingskorpus abgeglichen werden, in denen die vorkommenden Entitäten als solche und nach Art markiert und klassifiziert wurden (Schumacher, 2018a, §20). Das Problem dabei ist, dass selbst in großen Traingskorpora nie alle möglichen Bezeichnungen von Named Entities vorkommen können (Sutton and Mccallum, 2007, 5). Auch die Verwendung umfangreicher Listen (Gazetteers) z. B. gängiger Vornamen oder geographischer Orte hat sich als nicht erschöpfend erwiesen (Mikheev et al., 1999, 2). Darum muss die Software die Zugehörigkeit zu einer Objektklasse möglichst allein aus dem Kontext eines Wortes schließen (Sutton and Mccallum, 2007, 5). Der große Vorteil der kontextbezogenen Erkennung ist, dass gleiche Wörter ggf. unterschiedlichen Kategorien zugeordnet werden können (Mikheev et al., 1999, 1). So kann ein System durch Kontextinforma-

tionen z. B. die Stadt Paris von dem Paris der antiken Mythologie unterscheiden, der durch seine Liebe zur schönen Helena Troja ins Verderben stürzte. Die Erkennung kann entweder rein auf der Wortebene stattfinden oder es können auch die Objektklassen (Entity-Labels) der umliegenden Wörter mit einbezogen werden (Sutton and McCallum, 2007, 5). Dabei sind sowohl Positiv- als auch Negativwerte, also Entity-Labels ebenso wie die Kennzeichnung als nicht relevantes Objekt, für die Erkennung von Bedeutung. Es hat sich sogar gezeigt, dass es gewinnbringend sein kann, Kategorien in das Training einzubeziehen, die gar nicht vorkommen, da auch die Negativauszeichnung für diese Kategorien die Erkennungsgenauigkeit der anderen erhöhen kann (Sutton and McCallum, 2007, 23). Auf dieser Erkenntnis aufbauend wurden bei der vorliegenden Studie zunächst Raummetaphern in den Trainingsprozess einbezogen. Später hat sich gezeigt, dass es für diese Untersuchung kaum einen Vorteil gegenüber einem Kategoriensystem ohne Raummetaphern bietet, da – im Gegensatz zu einer nicht vorkommenden Kategorie – die Erkennungsgenauigkeit der Kategorie „Raummetapher“ in den unterschiedlichen Tests extrem stark schwankt und darum schwer kontrollierbar ist (vgl. Kapitel 6).

5.1.1 Conditional Random Fields (CRF)

Den Kern des hier angewendeten StanfordNER-Tools bildet eine Implementierung von Conditional-Random-Field-Algorithmen (CRF-Algorithmen). Dabei handelt es sich um vektorbasierte Vorhersagealgorithmen, die unterschiedliche, voneinander abhängige Variablen berücksichtigen können (Sutton and McCallum, 2010, 1). Die Basis bildet ein konditionales Sequenzmodell (Finkel et al., 2005, 3), in dem die einzelnen Objekte in Abhängigkeit zueinander gedacht werden. Zu den linearen und damit offensichtlichen Eigenschaften, wie z. B. der Position eines Wortes zwischen zwei anderen, werden weitere, versteckte Eigenschaften (Finkel et al., 2005, 3), wie z. B. die Zugehörigkeit zu einer Objektklasse, angenommen. Diese Eigenschaften sind beobachtbar (Finkel et al., 2005, 3) und können mit Hilfe von Hinweisen identifiziert werden. Zur Identifizierung werden Hinweiskombinationen mit Wahrscheinlichkeiten (Finkel et al., 2005, 3) verknüpft, die letztendlich die Vorhersage der Zugehörigkeit eines Wortes zu einer Objektklasse ermöglichen. Als Features werden die Merkmale bezeichnet, anhand derer Wörter klassifiziert werden können. Dies können orthografische Eigenschaften sein wie z. B. Präfixe und Suffixe oder auch die Position eines Wortes im Satzumfeld oder aber die Tatsache, ob dieses

Wort auf einer Liste vordefinierter Ausdrücke steht (Sutton and McCallum, 2010, 1). Die Schwierigkeit bei der technischen Umsetzung besteht darin, dass sich komplexe Feature-Kombinationen und gegenseitige Abhängigkeiten ergeben (Sutton and McCallum, 2010, 1). Dieser Schwierigkeit begegnen CRF-Implementierungen, indem sie zur Klassifizierung unterschiedliche graphische Modelle nutzen (Sutton and McCallum, 2010, 2). Dabei handelt es sich um eine Kombination linearer Modelle, die die Wort-Nachbarschaft berücksichtigen, wie z. B. das versteckte Markov-Modell oder das Modell linearer Ketten (Sutton and McCallum, 2010, 8–9 und 18), mit solchen, die zahlreiche Features eines Wortes mit einbeziehen, wie z. B. das Naive-Bayes-Modell (Sutton and McCallum, 2010, 10). Dabei können nicht nur Wörter im direkten Umfeld des jeweils betrachteten mit einbezogen werden, sondern auch solche, die an einer weit entfernten Position im Text stehen, die aber trotzdem Ähnlichkeiten zum betrachteten Wort zeigen (Sutton and McCallum, 2007, 2) nennen z. B. sog. skip-chain-Modelle und (Finkel et al., 2005, 1) Gibbs Sampling. Jedes Wort wird als Teil eines komplexen semantischen Feldes verstanden, in dem es Beziehungen zu anderen Wörtern im Text und zu (eigenen und fremden) Wort-Eigenschaften hat (Sutton and McCallum, 2010, 16).

5.1.2 NER als Preprocessing und als Unterstützung des hermeneutischen Prozesses

Wie in Abschnitt 5.1.1 bereits erwähnt, wird linguistische NER meist als Preprocessing-Methode klassifiziert, die bereits mit hoher Präzision mit Hilfe von out-of-the-box-Tools genutzt werden kann. Dass NER bereits auf einem hohen Niveau funktioniert, liegt daran, dass in unterschiedlichen Studien zahlreiche Features und Feature-Kombinationen getestet wurden. Bevor in Abschnitt 5.2 beschrieben wird, wie der Feature-Begriff in der NER-Forschung verstanden wird, welche Features getestet wurden, was derzeitig als State of the Art bezeichnet wird und was bei Adaptionen an die deutsche Sprache und die Domäne der Literatur berücksichtigt werden muss, soll hier ergänzend betrachtet werden, welche Chancen die Methode der digitalen Vorhersage birgt, wenn sie nicht nur als Preprocessing verstanden wird.

Die meisten Forschungslinien, in denen NER eine Rolle spielt, haben gemeinsam, dass sie die automatische Erkennung bestimmter sprachlicher Objektklassen optimieren. Das Ziel ist eine vollautomatisierte Annotation dieser Objektklassen in unbekannten Texten. Dies entspricht einem Preprocessing, das geisteswissenschaftlich Forschenden einen Analyseschritt komplett abnimmt. Während des Machine-Learning-Trainings in dieser Untersuchung zeigte sich allerdings, dass die Methode

auch anders fruchtbar gemacht werden kann.[3] Abgesehen davon, dass die CRF-Algorithmen nicht nur zur Annotation von Named Entities im engeren Sinne genutzt werden können, fiel auf, dass der Trainingsprozess an sich zur Schärfung des entwickelten Raummodells beitragen konnte. Die auf Basis der Forschungsliteratur entwickelten Kategorien von Raumdarstellungen können im hermeneutischen Sinne als Vorannahmen betrachtet werden. Diese werden bei der Annotation der Trainingskorpora auf den Text übertragen. Die Erkennungsraten geben Hinweise auf die Präzision der genutzten Kategorien, das Modell wird nachgeschärft, die Annotation der Trainingskorpora überarbeitet und das daraus entstehende CRF-Modell erneut getestet. Bessere oder schlechtere Erkennungsraten geben erneut Hinweise auf den Erfolg der Nachschärfung und das Modell kann bei Bedarf weiter überarbeitet werden. So entsteht ein iterativer Prozess, der dem hermeneutisch-zirkulären Prinzip stark ähnelt und der produktiv für den Erkenntnisgewinn genutzt werden kann. Die Methode wird hier also nicht nur zum Preprocessing genutzt, sondern als ein Kernbestandteil des Forschungsprozesses (ähnlich wie auch in Bögel et al., 2015, passim insbes. 27 und Gius und Jacke, 2015, passim beschrieben).

5.2 Features

Als Features werden charakteristische Eigenschaften von Wörtern verstanden, die so operationalisiert werden können, dass sie mithilfe von Algorithmen regelbasiert klassifiziert werden können (Sharnagat, 2014, 9–10). Insgesamt können vier Feature-Klassen unterschieden werden:

1. vektorbasierte Features (mathematische Abstraktionen des Texts),
2. wortbasierte Features (Worteigenschaften),
3. Listenfeatures (Verzeichnisse, Lexika),
4. Dokument- oder Korpusfeatures (Worthäufigkeiten, typischer Wortgebrauch) (Li et al., 2020, 4).

Bei der NER werden oft mehrere Regeln zusammen betrachtet und nur bestimmte Kombinationen von Eigenschaften führen letztendlich zur Annotation (Sharnagat, 2014, 9–10). Zwar wird durchaus darauf hingewiesen, dass sorgsame Feature-Auswahl für NER-Systeme, die überwachtes maschinelles Lernen nutzen, erfolgs-

[3] Diese Beobachtung stützt das von (Kuhn, 2018, 34) entworfene Szenario, dass experimentelle Vorhersagemodelle für die Überprüfung theoretischer Hypothesen genutzt werden können.

kritisch sein kann (Li et al., 2020, 4), bisher gibt es aber nur wenige Studien, die für die deutsche Sprache und die Domäne der deutschsprachigen Literatur Feature-Kombinationen testen (dazu gehört z. B. Jannidis et al., 2015). Für Kategorien des Raumes in literarischen Texten wurden solche Tests bisher nicht durchgeführt. Da eine technische Testreihe nicht im Fokus dieser Studie steht, kann dies auch hier nicht geleistet werden. Als Feature-Kombination wird darum die 15 Features umfassende Standard-Konfiguration des StanfordNER übernommen. Die genaue Feature-Kombination kann in der Properties-Datei (Raum.prop) im GitHub-Repository des Raum-Classifiers (Schumacher, 2021c) eingesehen werden. Eine Aufschlüsselung der Features findet sich in der StanfordNER-Feature-Factory unter https://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/ie/NERFeatureFactory.html [9.6.2021].

5.3 State of the art und Domänenadaption

Bei der Bewertung von Named-Entity-Recognition-Tools und -Modellen ist zu berücksichtigen, dass den Berechnungen von Erkennungsraten unterschiedliche Parameter zu Grunde liegen können (Marrero et al., 2013, 487–488). Diese können entweder auf den Annotationsprozess vor dem Training zurückzuführen sein, bei dem unterschiedliche Annotationsguidelines verwendet werden können (ebd.). Oder sie werden nach dem Trainingsprozess als Basis der Berechnung selbst festgelegt. Ein grundlegender Unterschied ist z. B., ob nur exakte Übereinstimmungen als korrekte Erkennungen gewertet werden oder ob auch teilweise Übereinstimmungen positiv bewertet werden (Marrero et al., 2013, 487–488; Li et al., 2020, 3–4).

Die Erkennungsgenauigkeit wird üblicherweise in Precision, Recall und F1-Score angegeben (Mikheev et al., 1999, 2). In Rückbezug auf (Sutton and McCallum, 2007, 28) wird hier folgende Token-basierte Variante der Formeln genutzt:

$$P(recision) = \frac{Anzahl\ der\ korrekt\ markierten\ Tokens}{Anzahl\ der\ markierten\ Tokens} \tag{5.1}$$

$$R(ecall) = \frac{Anzahl\ der\ korrekt\ markierten\ Tokens}{Anzahl\ aller\ der\ Kategorie\ zugehoerigen\ Tokens} \tag{5.2}$$

$$F1(-Score) = (2xPxR)/(P + R) \tag{5.3}$$

Die besten NER-Tools für die englische Sprache erreichen Werte von 97 % Precision und 96 % Recall (Marrero et al., 2013, 490–492). Die 16 Systeme, die beim Shared Task zur Entwicklung sprachunabhängiger NER bei der CoNLL03 (Conference on Computational Natural Language Learning 2003) entstanden sind, zeigen Durchschnittserkennungsraten zwischen 60 und 90 % (Marrero et al., 2013, 490–492). StanfordNER zeigte in mehreren Tests eine gute Performanz. Aber, wie Marrero et al. (2013) nachweisen, werden in Tests nicht unbedingt standardisierte Erfüllungskriterien berücksichtigt. In Bezug auf Content Validity, also die Frage, ob die für die Tests genutzten Einheiten die Domäne angemessen repräsentieren, halten sie fest, dass NER Tools suggerieren, dass sie für alle möglichen Kontexte genutzt werden können (Marrero et al., 2013). Trainiert und getestet wurden sie aber nur mit wenigen Kategorien und in einer einzigen Textdomäne (der journalistischen) (Marrero et al., 2013). Als schwächsten Punkt der NER bezeichnen die Autoren die externe Validität also das Kriterium der Übertragbarkeit auf andere Anwendungsbereiche (Marrero et al., 2013). Sie stellen fest, dass die Dokumente, die zur Entwicklung von NER Tools genutzt werden nicht in einem Prozess zusammengestellt wurden, der Zufälligkeit einbezieht (Marrero et al., 2013). Die NER Korpora sind oft klein und unausgewogen. Als Beispiel nennen Marrero et al. ein System von 2005, das auf einem Trainingskorpus von nicht mehr als 50.000 Tokens aus unterschiedlichen, aber allesamt journalistischen, Genres basiert. Sie berichten von Performanzverlusten von 20–40 % bei der Übertragung auf eine andere Domäne (Marrero et al., 2013, 490–492). Aber nicht nur die Texte, auch die Kategorien können stark nach Domäne variieren (Marrero et al., 2013, 489). Statt also Tools generisch zu nutzen, kann es von Vorteil sein, mit eigenen Modellen NER für vollkommen andere Kategorien zu verwenden (Marrero et al., 2013, 489). Dabei sollte allerdings berücksichtigt werden, dass bei Weitem nicht für jede Domäne bereits annotierte Trainingskorpora vorhanden sind und dass selbst annotierte Trainingsdaten häufig zu hochspezifischen Modellen führen, die nicht unbedingt nachgenutzt werden können (Marrero et al., 2013, 489).

NER für die deutsche Sprache

Die deutsche Sprache stand bisher vergleichsweise selten im Fokus von NER-Projekten (Weber and Pötzl, 2014, 158). Eines der ersten frei verfügbaren NER-Modelle für die deutsche Sprache wurde 2010 von Faruqui und Padó vorgestellt. Es wurde mit einem Trainingskorpus von rund 220.000 Tokens erstellt, was in etwa dem Umfang des für das englische Modell verwendeten Trainingskorpus entsprach. Faruqui und Padó nutzten das CoNLL 2003 Korpus, das aus insgesamt 553 Dokumenten besteht. Da es wie oben erwähnt etwa 220.000 Tokens umfasst (was sie als eher kleines Trainingskorpus einstufen), beinhaltet es ca. 400 Tokens / Dokument

im Trainingskorpus. Die Testsets umfassten 155 bzw. 201 Dokumente und waren etwa 55.000 Tokens lang, umfassten also etwa 350 bzw. 270 Tokens / Dokument (Faruqui and Padó, 2010, 1–2). Der Recall-Wert lag hier nur bei 64 %, was zeigt, dass sprachspezifische Besonderheiten wie z. B. morphologische Komplexität oder die Verwendung von Groß- und Kleinschreibung sich direkt auf die trainierten Modelle auswirken (vgl. Faruqui and Padó, 2010, 1; Weber and Pötzl, 2014, 158). Mit Hilfe von semantischer Generalisierung gelang es Faruqui und Padó, ihr Modell für die deutsche Sprache so weit zu verbessern, dass es am Ende eine Erkennungsrate von 79,8 % F1-Score erreichte (Faruqui and Padó, 2010, 1–2). Da der Vergleich mit anderen Modellen zeigt, dass nur ein weiteres System eine ähnlich gute Performanz erreicht (um 1 % besser bei der Kategorie „Person“), und zwar das auf dem Parser SynCoP beruhende System von Didakowski et al. (2007), legten diese Modelle für lange Zeit den State of the Art fest.

Im Jahr 2014 wurde mit GermEval eine Shared Task ausgerufen, die zum Ziel hatte, Named Entity Recognition für die deutsche Sprache weiter zu verbessern. Da das am besten abschneidende System hier mit eigenen Testdaten nur 73,26 % erreichte, mit den offiziellen GermEval-Testdaten sogar nur 54,55 % (Weber and Pötzl, 2014, 160–161), wurde allerdings keine Verbesserung gegenüber Faruqui und Padó oder Didakowski et al. erreicht. Erst neueste Studien zur Kombination von CRF-Implementierungen mit Word2Vec erhöhen den State of the Art auf eine Performanz von rund 85 % F1-Score (erreicht mit CoNLL-Vergleichsdaten aus der journalistischen Domäne) (Wiedemann et al., 2018, 6).

Bei der Durchsicht der unterschiedlichen Tooltests fallen zwei Dinge auf. Zunächst einmal hat NER für die deutsche Sprache ohne Word2Vec-Einbezug selten höhere Gesamtwerte erzielt als die Modelle von Faruqui und Padó und Didakowski et al. mit rund 80 % F1-Score. Die vergleichenden Tooltests zeigen aber häufig nicht nur die Gesamterkennungsraten, sondern auch die für einzelne Kategorien. Diese zeigen, dass die Kategorie „Person“ die besten Einzelerkennungsraten erreicht und häufig rund 10 % höhere F1-Scores erreicht als die anderen Kategorien (Faruqui and Padó, 2010, 4).

Für die vorliegende Studie sind neben den Versuchen der Adaption und Entwicklung von NER-Systemen für die deutsche Sprache vor allem auch die Anpassungsbemühungen an die Domäne der literarischen Texte von Bedeutung. Grundsätzlich sollte berücksichtigt werden, dass in der Computerlinguistik bereits etablierte Modelle vor dem Einsatz in der Literaturwissenschaft meist stark überarbeitet und angepasst werden müssen (Kuhn, 2018, 12). Dennoch wird NER durchaus auch mit Hilfe von off-the-shelf-Tools eingesetzt (vgl. Vala et al., 2015). Bisher wurde NER z. B. bereits recht häufig im Zusammenhang mit netzwerkanalytischen Ansätzen genutzt (Vala et al., 2015, 769–770). Wallace verwendete die Methode, um

sich narrativen Strukturen in einem Erzähltext zu nähern (vgl. Wallace, 2012). Auch im Zusammenhang mit der Analyse stereotyper Genderdarstellungen und -bewertungen hat sich die Methode als fruchtbar erwiesen, wenn auch hier nur in Kombination mit dem Training eines eigenen Modells (vgl. Schumacher und Flüh, 2020). Neben den eher inhaltlich ausgerichteten Analysen wurden auch bereits erste Trainingskorpora für literaturwissenschaftliche NER publiziert. Für englischsprachige Literatur wurde im Jahr 2019 ein Datenset mit sechs Entity-Kategorien bereitgestellt (vgl. Bamman et al., 2019). Für die deutsche Sprache wurde mit dem DROC (Deutsches Roman Korpus) 2017 annotiertes Trainingsmaterial publiziert, das derzeit für die NER allerdings nur Annotationen der Kategorie „Person“ bzw. „Figur“ aufweist (vgl. Horstmann, 2019) und keine für Raum.

Im Jahr 2007 stellten Borin et al. ein NER-System für schwedischsprachige Literatur des 19. Jahrhundert vor, das sehr gute Erkennungsraten von 89,6–92,8 % F1-Score erreichte (Borin et al., 2007, 7). Für die Kategorie „Location“ waren diese Werte mit 80,5–86,4 % etwas niedriger (Borin et al., 2007, 7). Interessant ist hier auch, dass dieses NER-System ohne Machine-Learning-Komponente auskommt und stattdessen mit einer selbst erstellten Grammatik operiert, die aus Identifikation von regelhaften Hinweisen auf Named Entities, Listenabgleich, Berücksichtigung semantischer Ähnlichkeiten, Informationsdisambiguation und Fehleranalyse beruht (Borin et al., 2007, 3–6). Hier werden außerdem recht viele Named-Entity-Kategorien einbezogen, nämlich acht, die außerdem mit Unterkategorien (insgesamt 57) versehen sind (Borin et al., 2007, 3). Mit LitNER wurde für die englische Literatur auch ein NER-System entwickelt, dass allerdings ausschließlich auf nicht-überwachten Machine-Learning-Techniken beruht (Brooke et al., 2016). Beide Ansätze sind auf die vorliegende Analyse nicht übertragbar, da hier gerade der Prozess des überwachten Lernens in den Mittelpunkt gerückt und für die Interpretation von Kategorien der Darstellung von Raum in der Literatur genutzt wird.

Im Jahre 2015 haben Jannidis et al. das Modell von Faruqui und Padó für deutschsprachige literarische Texte getestet und festgestellt, dass es lediglich einen F1-Score von 31 % erreicht Jannidis et al., 2015, 1–2). Damit ist eine Verringerung von rund 48 % gegenüber dem F1-Score der Ursprungsdomäne zu verzeichnen. Jannidis et al. haben, wie bereits erwähnt, daraufhin die Domänenadaption für die Kategorie der Figur in literarischen Texten vorangetrieben und unterschiedliche Feature-Kombinationen getestet (vgl. Jannidis et al., 2015). Basis der Tests (Trainingsdaten und Testdaten) waren das CoNLL Zeitungskorpus von 2003, ein Romankorpus bestehend aus 130 Sätzen aus 50 Romanen und einer Größe von 140.000 Tokens und einem Romankorpus aus 85 Romanen mit einem Umfang von 265.000 Tokens (Jannidis et al., 2015, 2–3). Für unüberwachte Lernverfahren wurde ein nicht-annotiertes Zeitungskorpus und ein nicht-annotiertes Romankorpus aus 60

Romanen verwendet. Es wurden sechs Basis-Features in unterschiedlichen Kombinationen mit sieben weiteren Features getestet. Die höchste Verbesserung ergab sich durch das Hinzufügen eines Word2Vec-Features (Jannidis et al., 2015, 2–3). Mit der besten Feature-Kombination wurde eine Precision von 94,86 %, ein Recall von 85,6 % und ein F1-Score von 89,98 % bei cross-Validierung (Jannidis et al., 2015, 2–3) und rund 83 % beim leave-one-out-Testverfahren (mehr über die unterschiedlichen Testverfahren in Abschnitt 6.1.2) erreicht (Jannidis et al., 2015, 5).

Damit gibt es ein performantes NER-System für deutschsprachige Erzählliteratur, das eine Kategorie abdeckt und den State of the Art auf 83 % bei leave-one-out-Testverfahren festlegt. NER kann damit allerdings nicht als gelöstes Problem für die Domäne der deutschsprachigen Literatur betrachtet werden. Da aber Systeme wie StanfordNER so konzipiert wurden, dass sie anpassbar sind, können sie zur Adaption weiterer Named-Entity-Kategorien genutzt werden. Dabei ist allerdings zu berücksichtigen, dass wahrscheinlich mittelgroße bis große annotierte Korpora erst noch erstellt werden müssen (Marrero et al., 2013, 494). Neben der Domänenadaption können Verfahren der intensiven Nachanalyse dabei helfen, die automatische Erkennung für die Literaturwissenschaft fruchtbar zu machen (Kuhn, 2018, 28).

5.4 Methodenkritik

Machine-Learning-Methoden können die literaturwissenschaftliche Annotation in mehrfacher Hinsicht voranbringen. Ein fertig trainiertes Tool kann zu enormer Zeitersparnis führen und ermöglicht es darum häufig, größere Textmengen zu berücksichtigen. Außerdem kann ein reflektierter Machine-Learning-Prozess sehr zur Erstellung, Überarbeitung und Schärfung von Kategorien und Kategoriensystemen beitragen. Aber die Methode weist auch einige Mängel auf, die nicht nur benannt sein sollen, sondern im Verlauf der hier dokumentierten Studie auch laufend mitgedacht und berücksichtigt werden.

5.4.1 Problematik der suggerierten Objektivität

NER ist eine Technik, die aus vielen bis sehr vielen Daten lernt, bestimmte Kategorien in Texten zu erkennen, die nicht im Trainingsmaterial enthalten sind und die auf Worteigenschaften basierende Features vergleicht und auswertet, um auf dieser Basis Vorhersagen darüber treffen zu können, welche Wörter in diese Kategorie fallen. Diese Charakteristik des Machine Learnings führt dazu, dass Implementierun-

gen dieser Art den Eindruck vermitteln, vom Menschen unabhängig, selbsttätig und damit rein maschinell und objektiv zu funktionieren. Aber dieser Eindruck täuscht. Zwar wird im bereits beschriebenen Operationalisierungsprozess tatsächlich der Versuch unternommen, Kategorien möglichst klar zu fassen und zu beschreiben, Features ganz klar zu definieren und damit die Variablen genau zu kontrollieren. Das ändert aber nichts daran, dass dieser Prozess in hohem Maße von Interpretationen abhängt. Die Auswahl der Trainingsdaten beruht ebenso auf Interpretation wie die der Kategorien. Auch was genau in diese Kategorien fällt und darum in der Vorbereitung des Trainingsprozesses auf bestimmte Weise annotiert wird, ist subjektiv und interpretationsbedingt. Dass in der Tat der gesamte Modellierungsprozess digitaler Studien ein subjektiver ist, stellt auch Piper seinen Distant-Reading-Analysen *Enumerations* voran (Piper, 2018, 11–12). Auf ähnliche Weise legt auch Dobson die Subjektivität des digital-geisteswissenschaftlichen Forschungsprozesses seinen Ausführungen in *Critical Digital Humanities* zu Grunde (Dobson, 2019, 20). Selbst wenn dieser Prozess intersubjektiv oder gar projektübergreifend durchgeführt wird (wie z. B. in der Shared Task von Willand, Gius und Reiter 2019), ist das Ergebnis nicht objektiv, sondern basiert auf Interpretationen. Die Beteiligung menschlicher Individuen am Trainingsprozess kann aber immerhin durch Kooperationen diversifiziert und somit ein Schritt in Richtung Überindividualität gemacht werden. Annotieren zwei oder mehr Personen ein Korpus, so kann die Übereinstimmung auf Grundlage eines Inter-Annotator-Agreements (mehr dazu in Abschnitt 6.3) berechnet und die Annotation weiter verfeinert werden.

In dieser Studie ist das allerdings aufgrund des Forschungssettings als Einzelstudie weitgehend unmöglich. Darum soll an dieser Stelle betont werden, dass die Methode des maschinellen Lernens nicht nur nicht objektiv ist, sondern, dass sie in dieser Arbeit höchst subjektiv eingesetzt wird. Da aber viele literaturwissenschaftliche Methoden interpretativ sind und trotzdem verwertbare und belastbare Erkenntnisse hervorbringen, wird die mangelnde Objektivität hier zwar berücksichtigt, sie führt aber nicht dazu, die bereits beschriebenen Vorteile der Methode auszuhebeln.

5.4.2 Problematik der unausgewogenen Ausgangsdaten

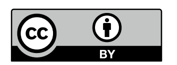
Eine ähnliche Problematik ist die mangelnde Repräsentativität der Datenbasis. Da hier mit literarhistorischen und zeitgenössischen Texten gearbeitet wird, die in digitaler Form vorliegen müssen, ist die Auswahl der Primärtexte von vornherein begrenzt. Bei den Texten aus dem 18.–20. Jahrhundert handelt es sich zum großen Teil um kanonisierte Werke. Es sind also Texte, die von Kritik, Wissenschaft, Literaturbetrieb, Bildungseinrichtungen und ähnlichen Instanzen als so bedeutsam

eingestuft wurden, dass sie ihren Weg in den literarischen Kanon gefunden haben. Diese Auswahl kann keinesfalls als repräsentativ für die Literaturproduktion einer bestimmten Zeit betrachtet werden. Hinzu kommt, dass die Datenbasis hier im Vergleich zu anderen Projekten, in denen Machine Learning zum Einsatz kommt, sehr klein ist. Hier kommen neben der begrenzten Textauswahl auch Aspekte der Performanz der eingesetzten Tools und Analysemethoden zum Tragen. Auch wenn hier das Kernkorpus nur 100 Texte umfasst, so sind die darin automatisch eingefügten Annotationen bereits so zahlreich, dass das eingesetzte Auswertungs- und Visualisierungstool CATMA an seine Grenzen stößt. Noch entscheidender ist aber, dass die Menge der Daten bereits so umfangreich ist, dass die Ergebnisse nur noch wenige besondere Phänomene zeigen. Da hier aber nicht nur Gemeinsamkeiten, sondern auch Unterschiede und Besonderheiten der vier Teilkorpora und deren literarhistorische Entwicklungen gezeigt werden sollen, hat sich die Größe der Datenbasis im Hinblick auf das Ziel dieser Studie als angemessen erwiesen. Dabei muss natürlich stets berücksichtigt werden, dass es sich nicht um eine repräsentative Auswahl handelt und dass sich in dieser Auswahl eine Reihe versteckter Vorannahmen verbirgt, die nicht alle bewusst gemacht und kontrolliert werden können. Im Hinblick auf die Zusammenstellung wurde durch die Wahl eines Zufallssystems allerdings der Versuch unternommen, zumindest an dieser Stelle keine weitere subjektive Verzerrung hinzuzufügen.

5.4.3 Problematik der Binarität

Ein in methodischer Hinsicht zentrales und grundlegendes Problem der literaturwissenschaftlichen Verwendung von Machine-Learning-Methoden und insbesondere der NER, ist, dass die hier verwendeten Kategorien binär im Hinblick auf die Zugehörigkeit angelegt sind. Das heißt, dass die Grundannahme lautet: Entweder etwas (ein Wort, eine Phrase, ein Absatz) fällt in eine der NER-Kategorien oder nicht. Nun sind aber literarische Texte bis hin zu ihren kleinsten Einheiten ambivalent oder sogar mehrdeutig. Dass ein literaturwissenschaftliches Phänomen eindeutig in eine Kategorie und nicht gleichzeitig noch mindestens in eine andere fällt, ist eher selten. In der hier dokumentierten Studie hat sich die Problematik der binären Kategorienzuordnung sehr deutlich gezeigt. Schon während des iterativen Trainingsprozesses des CRF-Modells, das im folgenden Kapitel beschrieben wird, traten häufig Momente ein, in denen ich die bisherige Kategorienzuordnung plötzlich hinterfragt und dann manchmal umorganisiert habe, sodass die Annotation von neuem beginnen musste. Am Ende war die sehr genaue Erstellung von Annotationsguidelines mit bestimmten Festsetzungen der hier gewählte Weg, um trotzdem zu einem relativ klar definier-

ten Kategoriensystem zu kommen. Doch auch wenn die Höherordnung einzelner Kategorien über andere und die damit verbundene Wenn-dann-Logik „wenn ein Wort in diese Kategorie fallen kann, ist das wichtiger als die zusätzliche Zugehörigkeit zu einer anderen“, eine praktikable Strategie ist, so löst sie die in den Texten, manchmal in einzelnen Wörtern, angelegten Mehrdeutigkeiten nicht auf. Um diese besser einschätzen zu können, wurde ein Test zum Inter-Annotator-Agreement mit zwei voneinander unabhängig annotierenden Literaturwissenschaftlerinnen[4] durchgeführt (vgl. Abschnitt 6.3). Auch bei dieser Problematik ist der einzig gangbare Weg die Reflexion und Bewusstmachung, dass am Ende von mit Machine Learning unterstützten literaturwissenschaftlichen Studien nur eine von mehreren möglichen Lesarten steht.



[4] Es handelt sich dabei um zwei studentische Hilfskräfte aus dem forTEXT-Projekt (siehe https://fortext.net), die beide bereits Erfahrung mit literaturwissenschaftlicher Annotation haben.

6 Maschinelles Erlernen der narrativen Raumindikatoren

Der Trainingsprozess des hier eingesetzten Machine-Learning-Tools StanfordNER wurde iterativ gestaltet. Bevor das Training begonnen werden konnte, mussten Vorarbeiten geleistet werden, wie z. B. die Zusammenstellung des Trainingskorpus, die im ersten Unterabschnitt dieses Kapitels 6.1 beschrieben werden. In insgesamt fünf Durchgängen habe ich die Trainingsdaten mit den Raumkategorien manuell annotiert, die Annotationen verbessert und die Kategorisierung hinterfragt und überarbeitet. Das Training wurde zunächst mit einem fünf Unterkategorien umfassenden Raummodell durchgeführt (erste Iteration). Anschließend wurde das Modell nachgeschärft und die Kategorienanzahl auf sieben erhöht (zweite Iteration). Das Trainingskorpus wurde in einer dritten Iteration erweitert, in einer vierten wurde die Konsistenz der Annotationen erhöht und die Annotationsguidelines überarbeitet. Schließlich wurde das Trainingskorpus erneut erweitert (fünfte Iteration). Ziel war es, ein CRF-Modell zu erstellen, dass es dem Tool ermöglicht eine durchschnittlich und über alle Kategorien hinweg berechnete mindestens 70 %-ige (F1-Score) Übereinstimmung mit manuellen Annotationen in Testtexten aus allen vier berücksichtigten Jahrhunderten, also dem 18.–21. Jahrhundert, zu erreichen. Ausführliche, auf GitHub dokumentierte Tool-Tests von Richter-Pechanski (2017) haben gezeigt, dass StanfordNER auf vortrainierten und eigens erstellten Daten, auf der Trainingsdomäne entsprechenden und domänenfremden Daten im Durchschnitt rund 68 % F1-Score erreicht. Im Vergleich mit anderen Tools in diesem Test erreicht StanfordNER mit 71,9 % außerdem den Höchstwert auf domänenfremden Daten. Der hier angestrebte F1-Score orientiert sich an diesen Ergebnissen. Gleichzeitig darf im

---

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-662-66035-5_6.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_6

Trainingsprozess nicht nur die Optimierung der Performanz des eingesetzten Tools im Fokus stehen. Darum wurden Anpassungen der Annotationskategorien stets nur in einem Umfang umgesetzt, der von der theoretischen Basis, die im ersten Teil dieses Buches im Zentrum stand, getragen werden kann. Nach jeder Iteration wurden die Resultate in einem Zwischenfazit beurteilt und das weitere Vorgehen entsprechend angepasst. Der letzte Abschnitt dieses Kapitels 6.2 dokumentiert Endresultate und abschließende Bewertung des entstandenen CRF-Modells.

6.1 Vorarbeiten

Auf Basis der bisher durchgeführten Domänenadaptionen für die deutsche Sprache im Allgemeinen (vgl. Didakowski et al., 2007; Faruqui and Padó, 2010) und für literarische Texte im Besonderen (vgl. Jannidis et al., 2015; Schumacher und Flüh, 2020), wurde hier aus einem Gesamtkorpus von rund 400 Romanen Untersuchungs-, Trainings-, und Testmaterial ausgewählt. Bei dem Gesamtkorpus handelt es sich um Texte aus dem TextGrid-Repository (TextGrid 2022) im txt-Format, die eine literarhistorische Zeitspanne vom 17.–20. Jahrhundert abdecken. Da die Anzahl der früheren Texte allerdings relativ gering war, wurde hier eine Einschränkung auf die Zeit vom 18.–20. Jahrhundert vorgenommen. Die Textsammlung habe ich mit selbst gescannten oder als Ebook erworbenen Romane aus dem 21. Jahrhundert ergänzt. Im Folgenden wird kurz dokumentiert, wie aus diesem Gesamtkorpus die betrachtete Teilmenge ausgewählt wurde und welche Verfahren in dieser Studie zum Testen der Modelle angewendet wurden.

6.1.1 Korpusgestaltung

Aus dem Gesamtkorpus wurden zuerst vier Teilkorpora gebildet. Für das 18., 19., 20. und 21. Jahrhundert wurden jeweils 25 Texte ausgewählt, sodass das Untersuchungskorpus insgesamt 100 Texte umfasst. Die Texte wurden per Zufallsgenerator aus dem Gesamtkorpus ausgesucht, damit persönliche Präferenz nicht in die Datenbasis einfließen und diese zusätzlich verzerren kann. Die Texte des TextGrid-Repositories sind in guter Qualität aufbereitet, es handelt sich aber hauptsächlich um kanonisierte Literatur (Horstmann und Kern, 2018), sodass, aufgrund der vom Archiv durchgeführten Vorauswahl, keine objektive Datenbasis erreicht werden kann. Auch ein Anspruch auf Repräsentativität für die Literaturproduktion der abgedeckten Zeitspannen kann nicht erhoben werden.

An die Beobachtungen von Dobson (Dobson, 2019, 66–100) zur Historizität digitaler Texte anknüpfend, wurden die Trainingsdaten zeitspezifisch, d. h. jeweils den Jahrhunderten 18, 19, 20 und 21 entsprechend, aufbereitet. Den Ausführungen von Jannidis et al. (2015) folgend wurden als Ausgangspunkt Teilkorpora von jeweils 40.000 Tokens angelegt. Im Vergleich zu den Testreihen von Jannidis et al. (2015) wurden allerdings relativ lange Textpassagen und keine zufälligen, sondern jeweils die Anfänge aus den einzelnen Romanen extrahiert. Dieses Vorgehen hat sich in eigenen, vorbereitenden Tests gegenüber einer Zusammenstellung aus kürzeren, zufällig ausgewählten Textabschnitten bewährt (Schumacher, 2020a). Dass sich gezeigt hat, dass längere Passagen von Romananfängen im Trainingskorpus die Erkennungsquoten der automatischen Annotation von Ausdrücken, die zur Raumdarstellung genutzt werden, positiv beeinflusst, ist durch die expositorische Funktion derselben begründet. So ist durch zahlreiche Studien belegt, dass am Romananfang stets eine Art von Einführung der Lesenden in die Textwelt oder auch ein Aufbau derselben steht (vgl. z. B. Miller, 1965, 9; Piwitt, 1965, 173; Thüsen, 1969, 623; Erlebach, 1990, 39 ff.; Retsch, 2000, 138; Richardson, 2008, 4; Bieker, 2019, 37 ff.). Retsch bezeichnet die Eingangssequenz von Romanen, vor allem derjenigen aus dem 19. Jahrhundert, sogar als „vergleichbar mit einer Beschreibung in einem Reiseführer“ (Retsch, 2000, 140). Für jedes der vier Teiltrainingskorpora wurden zuerst zehn Texte per Zufallsgenerator ausgewählt. Die einzige Einflussnahme betraf das Autorengender. Da insgesamt sehr viel mehr männliche Autoren im Literaturkanon vertreten sind und die Textdaten des Gesamtkorpus zu einem großen Teil von Schriftstellern verfasst sind, wurde hier der Versuch unternommen dieses Missverhältnis auszugleichen. Die Trainingskorpora wurden also zu gleichen Anteilen aus Texten von Schriftstellerinnen und von Schriftstellern gebildet. Dabei ist mir bewusst, dass Autorengender nur eine von zahlreichen Variablen ist, die auf den Kanonbias einwirken. Mich interessiert hier aber besonders die statistische Verzerrung durch die Unterrepräsentation von Autorinnen, da Raumdarstellungen in der feministischen Narratologie bereits mit der Gestaltung von Figurengender in Zusammenhang gebracht wurde (vgl. Abschnitt 2.2.1) und dieser Aspekt darum für Anschlussuntersuchungen und auch für die Nachnutzung des Raum-Classifiers besonders vielversprechend ist. So könnte z. B. untersucht werden, ob sich die figurengenderspezifische Raumdarstellung von Autoren und Autorinnen unterscheidet, ohne dass sich dabei lediglich eine, durch das Trainingskorpus bedingte, schlechtere Erkennung der Raumdarstellung von Schriftstellerinnen störend niederschlagen könnte. Dass es nicht möglich ist, in Daten, die als Grundlage von Machine-Learning-Prozessen genutzt werden, versteckte Vorannahmen und gesellschaftliche Ungleichheiten vollends auszugleichen, zeigen Caliskan, Bryson und Narayanan (2017). Zur Problematik der Unausgewogenheit digitaler Archive haben sich z. B.

Risam (Risam, 2019, 67) und Dobson (Dobson, 2019, 69) geäußert. Dobson geht dabei auch auf die Verzerrung von Ergebnissen in Natural-Language-Processing-Analysen durch unpassenden Datenabgleich ein (Dobson, 2019, 86–93).

Nachdem die ersten Versuche mit den jeweils 40.000 Tokens umfassenden Teilkorpora durchgeführt worden waren, wurden die Trainingsdaten schrittweise um jeweils 8.000 Tokens erweitert, um das CRF-Modell zu verbessern. Als Testdaten wurden jeweils zwei Texte pro Jahrhundert aus dem Untersuchungskorpus ausgewählt. Einer dieser Testtexte ist jeweils von einem Autor, der andere von einer Autorin. Die Testtexte unterscheiden sich in ihrer Länge, sodass dieser Faktor bei den Testauswertungen berücksichtigt werden kann. Eine Auflistung der Texte, die im Untersuchungs-, Trainings und Testkorpus enthalten sind, kann im elektronischen Zusatzmaterial eingesehen werden.[1]

6.1.2 Auswahl der Test-Methode

In den Testreihen von Jannidis et al. (2015) hat sich gezeigt, dass sich die Wahl des statistischen Auswertungsverfahrens entscheidend auf die Testergebnisse auswirkt. Beim strengeren *leave one out*-Verfahren werden die Testdaten komplett aus dem Trainingsprozess ausgeschlossen. Die Tests laufen also auf einem oder mehreren Texten oder Testpassagen, die dem Classifier in Gänze unbekannt sind. Bei der *cross validation*-Methode wird dagegen das Trainingskorpus in eine vorher festgelegte Anzahl gleich langer Abschnitte unterteilt. Reihum wird jeder dieser Abschnitte ein Mal aus dem Trainingsprozess herausgehalten und das entstandene Modell dann auf dem ausgeschlossenen Abschnitt getestet. Am Ende werden die Ergebnisse aller Tests gemittelt. Wird das Trainingskorpus in zehn Teile unterteilt spricht man z. B. von *ten fold cross validation*. Da die Abschnitte in gleich lange Passagen unterteilt werden, kann es dabei sein, dass ein Teil eines Textes im Trainingskorpus erhalten bleibt, während der andere sich im Testabschnitt befindet. Die Testdaten können also teilweise im Trainingskorpus vorhanden sein. Ein Beispiel: Im ersten Teil eines Textes (Ausschnitt im Trainingskorpus) häufig erwähnte Ortsnamen könnten dann auch im zweiten Teil desselben (Ausschnitt im Testkorpus) häufig sein und darum beim Test besser erkannt werden. Darum führt *cross validation* häufig zu besseren Testergebnissen. In den Tests der vorliegenden Studie wurde das strengere leave-one-out-Verfahren genutzt. Abweichend vom in der NER weit

[1] Trainings- und Testdaten wurden soweit rechtlich möglich zur Prüfung und Nachnutzung im GitHub-Repository des Raum-Classifiers veröffentlicht (Schumacher, 2021a). Auflistungen der genutzten Romane und ausführliche Testdaten sind dort ebenfalls einsehbar.

verbreiteten Verfahren, ein Referenzkorpus zu erstellen (Kuhn, 2018, 14), das in seiner Gestaltung dem Trainingskorpus entspricht (also in diesem Fall aus 4.000-Token Passagen mehrerer Romane zusammengesetzt worden wäre) wurden hier erst ganze Romane, später 10.000-Token-Abschnitte als Testdaten aufbereitet (mehr zu den Test-Iterationen in Abschnitt 6.2). Auf diese Weise kann nämlich auch betrachtet werden, ob im Verlauf der Erzählung die Erkennungsraten konstant bleiben. Im Test wurde also jeweils die automatische Annotation mit der manuellen Annotation eines Erzähltextes, aus dem keine Passage in den Trainingsdaten enthalten ist, verglichen. Auf Basis dieses Vergleichs wurden vom StanfordNER Precision-, Recall- und F1-Werte für jede Unterkategorie und für die Oberkategorie Raum insgesamt berechnet. Um die umfangreichen und sehr ausführlichen Tests nachvollziehbar zu halten, wurden nur die F1-Score-Werte aufgelistet und zum Vergleich heran gezogen. Die F1-Score-Werte wurden jeweils mit dem anderen Vergleichstext für das jeweilige Jahrhundert und mit den Testtexten der anderen Jahrhunderte abgeglichen und es wurden Mittelwerte errechnet. Auf diese Weise wurde das Training sehr genau beobachtet und ggf. angepasst.

6.2 Trainingsprozess

Die Trainings- und Testdaten wurden manuell mit den Raumkategorien des oben umrissenen Modells annotiert. Da die gängige Praxis der Erstellung von hochqualitativen Gold-Standard-Daten (Wissler et al., 2014, Abschnitt III) in dieser einzeln durchgeführten Studie nicht übernommen werden konnte, wurde ein iterativer Prozess gewählt. Dieser Prozess wurde analog zum Prozess der Erstellung von Annotationsguidelines gestaltet, wie er von Reiter (2020, 193) beschrieben wird. Insgesamt wurden 5 Iterationen durchlaufen, die im Folgenden dokumentiert werden:

1. Training und Test eines Modells mit 5 Raum-Kategorien
2. Training und Test eines Modells mit 7 Raum-Kategorien
3. Ausweitung des Trainingskorpus von 160.000 auf 240.000 Tokens
4. Verbesserung des Trainingskorpus und der Testtexte durch Schärfung der Kategorien (Erstellen detaillierter Guidelines) und Erhöhung der Annotationskonsistenz
5. Ausweitung des Trainingskorpus von 240.000 auf 320.000 Tokens

Die Trainings- und Testdaten wurden in den Iterationen eins und zwei jeweils zwei Mal mit zeitlichem Abstand annotiert. Der zweite Annotationsdurchgang diente

der Überarbeitung und Verbesserung bzw. Anpassung an die Ausdifferenzierungen der Kategorien, die sich während des Annotationsprozesses ergeben haben. Um diesen langwierigen Prozess etwas zu beschleunigen, wurden die Daten mit dem StanfordNER vorannotiert, sobald die ersten Modelle trainiert waren. Dies erleichterte nicht nur die doppelte manuelle Annotation, sondern trug auch zu einer höheren Datenkonsistenz bei. Bei der literaturwissenschaftlichen Annotation gibt es oft Ambiguitäten und unauflösbare Mehrdeutigkeiten, die nicht nur dazu führen können, dass mehrere Annotierende zu unterschiedlichen Interpretationen kommen (Gius and Jacke, 2017, 246–248), sondern auch – und das zeigt der iterative Prozess des Machine-Learning-Trainings in dieser Studie ebenfalls – dass eine literaturwissenschaftlich annotierende Person sich im Verlauf einer langandauernden Annotation mal für eine Wortbedeutung und mal für eine andere entscheiden kann. Die automatische Vorannotation kann einen Hinweis darauf geben, wie im bisherigen Annotationsverlauf mit solchen Ambiguitäten umgegangen wurde und somit die Konsistenz in den Annotationsdaten erhöhen. In der ersten, zweiten und dritten Iteration wurde zunächst relativ zügig und lediglich mit Hilfe der Modellierung narrativer Raumindikatoren in Kapitel 2 und einer Skizze der Annotationsguidelines annotiert. Die in Kapitel 4 enthaltenen, ausführlichen Guidelines mit Definitionen und Beispielen für die einzelnen Kategorien, die sich aus der mit dem Trainingsprozess einhergehenden Schärfung derselben ergaben, wurden dann erst in der vierten Iteration festgehalten. Dazu gehört auch der Abgleich und zum Teil die Abgrenzung mit bzw. von den literaturwissenschaftlichen Ansätzen, die im ersten Teil dieser Arbeit zusammengefasst wurden.

6.2.1 Die erste Iteration

Abweichend vom oben beschriebenen Setting wurde die erste Testreihe nur mit jeweils einem Testtext pro Jahrhundert durchgeführt. Diese erste Testreihe sollte lediglich dazu dienen, zu prüfen, ob die Methode Anfangswerte der Erkennungsgenauigkeit zeigt, die darauf hindeuten, dass sie für diese Studie gewinnbringend eingesetzt werden kann. Darüber hinaus sollte das, zu diesem Zeitpunkt noch fünf Kategorien umfassende, Raummodell einem ersten Test unterzogen werden.

In der ersten Testreihe wurde jeder Testtext zuerst mit jedem der mit jahrhundertspezifischen Trainingsdaten erstellten CRF-Modelle getestet. Anschließend wurden die Trainingsdaten kumuliert, sodass neben den vier CRF-Modellen, die mit 40.000 Tokens umfassenden, jahrhundertspezifischen Trainingskorpora trainiert wurden, drei weitere entstanden. Eines davon umfasst 80.000 Tokens (18. und 19. Jahrhundert), ein zweites 120.000 Tokens (18., 19. und 20. Jahrhundert) und das dritte

160.000 Tokens (aus allen vier Jahrhunderten). Die Testtexte wurden dann auch mit diesen drei ergänzenden CRF-Modellen einzeln getestet, woraufhin die Durchschnittswerte pro Kategorie ermittelt wurden. Eine Auflistung der gesamten Testergebnisse befindet sich im GitHub-Repository zum Raum-Classifier (Schumacher, 2021a).

Insgesamt waren die Ergebnisse der ersten Testreihe noch nicht zufriedenstellend, wenn sie auch bereits zeigten, dass die Methode durchaus zu dem theoriebasierten Modell literarischen Raumes passt. Die Gesamterkennungsraten bei jahrhundertspezifischen Modellen, die mit jeweils 40.000 Tokens umfassenden Trainingskorpora trainiert wurden, liegen zwischen rund 37 % und rund 58 %. Mit den kumulativ erweiterten Trainingsdaten konnten diese Werte um gut 10 % auf rund 57–65 % verbessert werden. Die schnelle Steigerung der Erkennungsgenauigkeit durch Ausweitung des Trainingskorpus zeigt, dass die Methode weiteres Potential zur automatischen Erkennung von Indikatoren narrativer Raumdarstellung in Erzählungen bietet.

Jahrhundertspezifische Testreihe

Die ersten Tests der Testreihe wurden jeweils mit einem jahrhundertspezifischen Testtext und allen jahrhundertspezifischen CRF-Modellen durchgeführt. Für das 18. Jahrhundert wurde Schillers *Der Geisterseher* ausgewählt, der Testtext für das 19. Jahrhundert war Heydens *Der graue John*[2], für das 20. Jahrhundert fiel die Wahl auf Reventlows *Der Selbstmordverein* und der Testtext aus dem 21. Jahrhundert war Stanišićs *Vor dem Fest*. Jeder der annotierten Testtexte wurde mit vier CRF-Modellen getestet, die auf jeweils 40.000 Tokens aus zehn Erzähltexten des 18.,19.,20. und 21. Jahrhunderts trainiert wurden. Ausgewertet wurden Annotationen der Kategorien „Ort“, „Relation“, „Thema“, „Hinweis“ und „Metapher“ sowie alle Annotationen der Raumkategorien insgesamt.

Die Zahlenwerte der Gesamterkennungsraten schwanken je nach Testtext unterschiedlich stark. Während z. B. bei Schillers *Geisterseher* Quoten zwischen 43,16 % (mit Trainingsdaten aus dem 19. Jahrhundert) und 47,45 % (mit Trainingsdaten aus dem 20. Jahrhundert) erreicht werden, also nur eine Schwankung von rund 4 % vorliegt, liegen die Extremwerte für Stanišićs *Vor dem Fest* zwischen 38,47 % (Trainingsdaten aus dem 18. Jahrhundert) und 55 % (Trainingsdaten aus dem 21. Jahrhundert), und schwanken somit um rund 16,5 %. Die Tests zeigen sehr deutlich,

[2] Dieser Text wurde in den Tests ab der zweiten Iteration aber ausgetauscht und durch Andreas-Salomés *Ruth* ersetzt, um in den Testdaten ein ausgewogenes Verhältnis des Autorengender zu erreichen.

dass nicht immer die besten Erkennungsquoten mit Trainingsmaterialien erreicht werden, die aus demselben Jahrhundert stammen, aus dem auch der Testtext kommt.

Insgesamt liegen die F1-Score-Werte zwischen 38,47 % (bei Stanišić / Trainingsdaten aus dem 18. Jahrhundert) und 57,93 % (erreicht bei Reventlow / Trainingsdaten aus dem 21. Jahrhundert). Die Spanne zwischen der schlechtesten und der besten Erkennung umfasst also rund 19,5 %. Bei dreien der vier Testtexte ist die Erkennungsrate bei Verwendung des CRF-Modells am besten, das mit Daten aus dem 21. Jahrhundert trainiert wurde (vgl. dazu auch Abbildung 6.1). Ein Blick auf die Annotationsdaten zeigt, dass dies möglicherweise darauf zurückzuführen ist, dass im Trainingskorpus des 21. Jahrhunderts quantitativ die meisten Ausdrücke als zu einer der Raumkategorien zugehörigen Wörter und Phrasen annotiert wurden. Eine Ausnahme bildet Schillers *Geisterseher*, welcher die beste Erkennungsrate mit dem CRF-Modell des 20. Jahrhunderts erreicht (vgl. Abbildung 6.1). Hier wird deutlich, dass sorgfältig geprüft werden muss, warum ein CRF-Modell besser funktioniert als ein anderes. Ein voreiliger Schluss auf die inhaltliche Ebene, wie hier z. B., dass Trainingsdaten aus der gleichen Zeitspanne stammen müssen wie der zu annotierende Text, hält den hier durchgeführten Testreihen nicht stand.

Bei näherer Betrachtung zeigen die ersten Tests, dass die Kategorie „Ort" meistens am besten erkannt wird (vgl. Abbildung 6.1). Diese Beobachtung deckt sich mit der Annahme, dass diese eine der konkretesten Raumkategorien ist (siehe Abschnitt 4.1). Die einzige Ausnahme bildet hier der Testtext aus dem 21. Jahrhundert, Saša Stanišićs *Vor dem Fest* (Abbildung 6.1). Hier werden mit fast allen CRF-Modellen (die Ausnahme bildet das CRF-Modell des 18. Jahrhunderts) die besten Werte für die Kategorie „Relation" erreicht. Diese Kategorie wurde bei der Modellentwicklung ebenfalls als sehr explizit ausgemacht, schnitt aber in den anderen drei Testtexten etwas schlechter ab. Die beiden eher impliziten Raumkategorien „Raumthema" und „Raumhinweis" wurden schlechter erkannt, was allerdings auch wenig überrascht. Bei den beiden letztgenannten Kategorien ist die Beobachtung interessant, dass je nach Testtext entweder „Raumhinweise" oder „Raumthemen" besser erkannt wurden (vgl. Abbildung 6.1). Nur bei Reventlows *Selbstmordverein* gibt es sogar innerhalb der textspezifischen Testreihe einen Wechsel. Mit den ersten drei CRF-Modellen wurden „Raumhinweise" besser erkannt. Kombiniert man den aus dem 20. Jahrhundert stammenden Text mit einem CRF-Modell, das mit Trainingsdaten aus dem 21. Jahrhundert trainiert wurde, so werden „Raumthemen" besser erkannt. Die Metaphernerkennung zeigt durchgehend schlechte Ergebnisse, was nicht verwunderlich ist, hier aber auch nicht weiter hervorgehoben wird, da diese Kategorie nicht im Zentrum des Interesses steht.

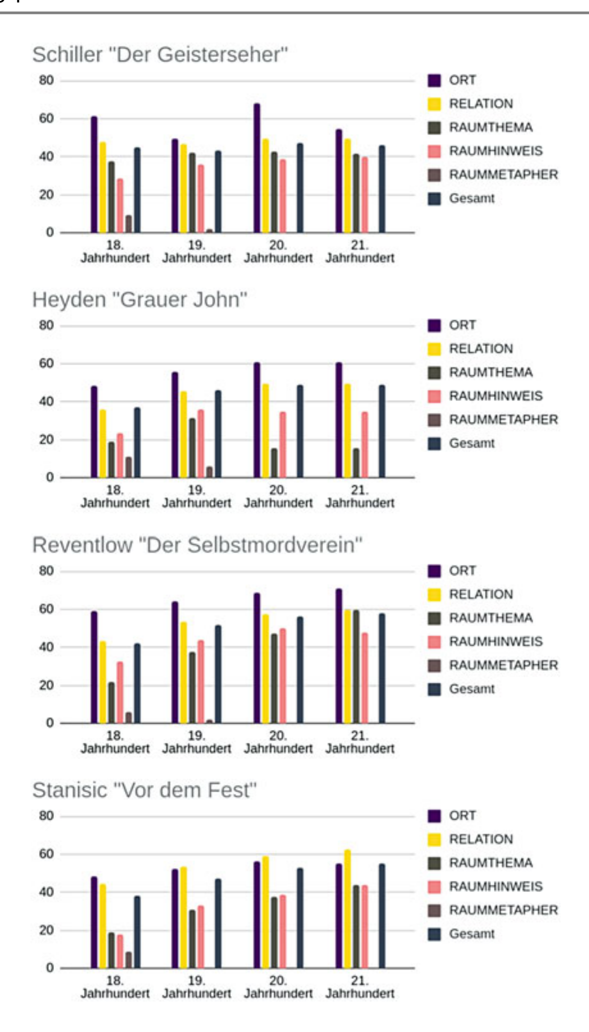


Abbildung 6.1 Visualisierung der Testergebnisse des CRF-Modells mit 5 Raumkategorien

Testreihe mit kumulierten Daten
Für den zweiten Test wurden die Ergebnisse der einzelnen Testtexte mit vergrößertem Trainingskorpus ermittelt. Für die Baseline-Werte des mit nur 40.000 Tokens trainierten Modells wurden für jeden Testtext die Ergebnisse mit dem Modell genommen, das mit jahrhundertspezifischen Trainingsdaten trainiert wurde, obwohl das nicht bei allen Tests die Modelle waren, die die besten Ergebnisse erzielt hatten.

Hier ging es darum, zu vermeiden, die jeweils besten Ergebnisse herauszupicken und stattdessen möglichst ähnliche Ausgangsbedingungen für die Tests zu schaffen. Im ersten Schritt wurden also die Gesamterkennungsquoten des Testtextes aus dem 18. Jahrhundert mit dem CRF-Modell des 18. Jahrhunderts, des Testtextes aus dem 19. Jahrhundert mit dem CRF-Modell des 19. Jahrhunderts usw. für jede Kategorie errechnet. Anschließend wurden die Mittelwerte der F1-Scores der Texte pro Kategorie für das CRF-Modell berechnet, das mit einem Trainingskorpus aus 80.000 Tokens aus Texten des 18. und 19. Jahrhunderts trainiert wurde. Daraufhin wurden dann die Erkennungsraten, sowohl für das aus den drei Jahrhunderten 18–20 und aus 120.000 Tokens bestehende CRF-Modell als auch für das 160.000 Tokens umfassende und vier Jahrhunderte beinhaltende CRF-Modell, ermittelt. Auf diese Weise entsteht ein diachroner Blick auf die Testdaten.

Um die Daten der ersten Testreihe besser einordnen zu können, lohnt sich zusätzlich ein Blick auf die quantitativen Besonderheiten der Trainingsannotationen.[3] In den jahrhundertspezifischen Trainingskorpora sind die Gesamtzahlen der annotierten Tokens der Kategorie „Ort“ vergleichsweise homogen und liegen zwischen 860 und 1.113 annotierten Tokens im jeweiligen Trainingskorpus. In den anderen Kategorien gibt es deutlich mehr Schwankungen. Die größte Variation gibt es in der Kategorie „Relation“, in der die annotierten Tokens zwischen 1.012 und 2.453 liegen. Insgesamt fällt auch auf, dass es wesentlich mehr Annotationen für die Kategorien „Ort“ und „Relation“ gibt als bei den anderen Kategorien. Der quantitative Vergleich der manuellen Annotationen mit den Testergebnissen zeigt, dass die vergleichsweise stark ansteigende Erkennung von Indikatoren für „Raumthemen“, bei der Verwendung der kumulierten Trainingsdaten, darauf zurück zu führen sein könnte, dass im Trainingskorpus des 21. Jahrhundert mit Abstand am meisten Ausdrücke als zu einem „Raumthema“ gehörig annotiert wurden. Überraschend scheint zunächst, dass die Anzahl der Annotationen für „Relationen“ und „Raumhinweise“ in den kumulierten Trainingsdaten zwar stark ansteigt (Abbildung 6.2 rechts), die durchschnittliche Erkennungsgenauigkeit der Testtexte sich für diese Kategorien aber vergleichsweise wenig verbessert (Abbildung 6.2 links). Während die Erkennungsgenauigkeit für Orte und Themen also mehr oder weniger konstant ansteigt, stagniert die Verbesserung des CRF-Modells für Relationen und Raumhinweise:

Aus diesem Befund habe ich geschlussfolgert, dass die Kategorien „Relation“ und „Raumhinweis“ im Raummodell noch zu unscharf umrissen sind. Für die nächste Iteration des Trainingsprozesses wurden diese beiden Kategorien deshalb jeweils in zwei Kategorien aufgespalten.

[3] Ausführliche Auflistung der Anzahl von Annotationen in den wachsenden Trainingsdaten finden sich im GitHub-Repository des Raum-Classifiers (Schumacher, 2021c).

Liegt der Fokus auf den Jahrhunderten und nicht auf den einzelnen Kategorien von Annotationen in den Trainingskorpora, so fällt auf, dass im Trainingskorpus des 21. Jahrhunderts mit Abstand die meisten expliziten und impliziten Raumausdrücke annotiert wurden und dass im Trainingskorpus des 20. Jahrhunderts die wenigsten Ausdrücke als Indikatoren für narrative Raumdarstellung klassifiziert wurden. Die Zahlen für das 18. und 19. Jahrhundert sind vergleichsweise konstant (vgl. Abbildung 6.3):

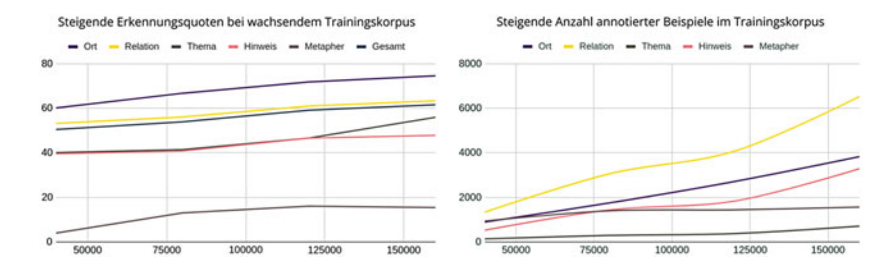


Abbildung 6.2 Vergleich der durchschnittlichen Erkennungsquoten pro Kategorie bei kumulierten Trainingsdaten mit der Anzahl der Annotationen pro Kategorie im wachsenden Trainingskorpus

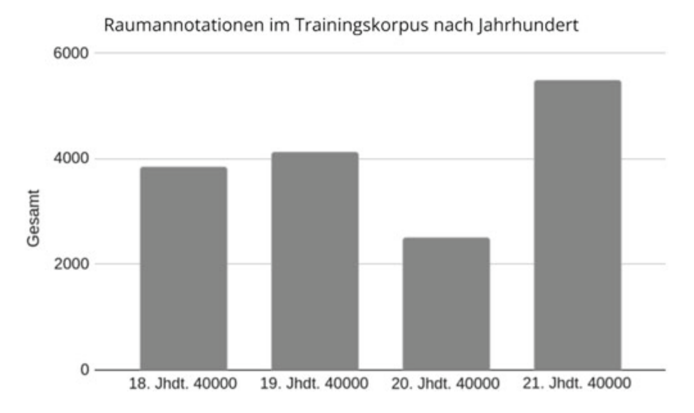


Abbildung 6.3 Anzahl der als explizite oder implizite Raumausdrücke annotierten Tokens pro Trainingskorpus

Wie oben bereits erwähnt, scheint sich zunächst die Erkennungsgenauigkeit damit zu decken, dass Raum mehr oder weniger explizit dargestellt werden kann. Die als expliziter eingeordneten Kategorien werden deutlich besser erkannt. Der Abgleich mit den Annotationsdaten konnte hier aber einen vorschnellen Kausalschluss verhindern. Der Erfolg des Modells in Bezug auf die einzelnen Kategorien hängt sehr viel wahrscheinlicher von zwei Faktoren ab, nämlich von der Anzahl der Beispiele, die in den Trainingsdaten vorkommen (vgl. Abbildung 6.3) und davon, wie eindeutig die jeweilige Kategorie spezifiziert wurde. Darum wurden in den folgenden Iterationen zunächst die Kategorien geschärft und ergänzt und dann das Trainingskorpus weiter ausgeweitet.

Resultate für das weitere Vorgehen
Vor allem die Kategorien „Relation“ und „Raumhinweis“ müssen klarer gefasst werden. Ein Blick auf die Trainingsdaten verrät, dass die hier annotierten Ausdrücke tatsächlich in sehr unterschiedliche Kategorien fallen. Zwei Beispiele aus den Trainingsdaten, die als zur Kategorie „Relation“ gehörig annotiert wurden, belegen diese Deutung. Das Wort „hinzukommen“ enthält z. B. die Relation „hinzu“ gehört aber zur Wortart der Verben. Dagegen handelt es sich z. B. bei „zwischen“ um eine deiktische Präposition, also um einen eindeutig relationalen Ausdruck. Im ersten Raummodell fallen diese auf linguistischer Ebene sehr unterschiedlichen Ausdrücke beide in die Kategorie „Relation“. Während die deiktische Präposition genau in diese Kategorie zu passen scheint, sollte für das Verb „hinzukommen“ eine neue Kategorie gefunden werden. Diese Kategorie habe ich „relationale Verben“ genannt. Mit einer solchen Aufspaltung der Kategorie geht allerdings auch einher, dass das gesamte Trainingskorpus erneut geprüft werden muss. Wenn nun nämlich Verben wie „hinzukommen“ in eine eigene Kategorie fallen, statt nur genau dann annotiert zu werden, wenn sie in dieser Form vorkommen, da sie nur dann als Präfix die Relation „hinzu“ enthalten, so sollte diese Kategorie auch alle Verben enthalten, die als relational interpretiert werden können. Das Verb „kommen“ muss dann z. B. immer annotiert werden, das Verb „gehen“ ebenfalls usw. Auf Basis solcher Detailanalysen wurde zunächst das Raummodell und dann, durch erneute Annotation, das CRF-Modell weiter geschärft.

Darüber hinaus schien eine Erweiterung des Trainingskorpus sinnvoll. Die Steigerung der durchschnittlichen Gesamterkennungsraten der Modelle mit kumulierten Daten um rund 10 % von 50,69 auf 61 % ist zwar insgesamt noch nicht hervorragend, zeigt aber, dass eine Ausweitung des Trainingsmaterials durchaus gewinnbringend ist. Die Testreihe zeigt, dass die Erkennung mit jahrhundertspezifischem CRF-Modell nicht bei allen Testtexten besonders hohe Erkennungsraten erreicht. Die Erweiterung kann also auch weiterhin kumulativ erfolgen, d. h. die Trainingskor-

pora wurden in der nächsten Iteration um 20.000 Tokens pro Jahrhundert erweitert und dann wurden die Trainingsdaten alle zusammengefasst, sodass am Ende ein übergreifendes Trainingskorpus von 240.000 Tokens entstand.

6.2.2 Die zweite Iteration

Der entscheidende Unterschied zwischen der ersten und der zweiten Iteration ist die Ausweitung des Raummodells auf sieben Kategorien. In der hier dokumentierten Testreihe steht vor allem der Vergleich dieses sieben Kategorien umfassenden Modells mit dem fünf-Kategorien-Modell aus der ersten Iteration im Vordergrund. Das Test-Setting ähnelt dem der ersten Iteration – auch hier wurde zunächst mit jahrhundertspezifischen CRF-Modellen getestet, die mit jeweils 40.000 Tokens umfassenden Trainingskorpora trainiert wurden. Anschließend wurden die Trainingsdaten kumuliert. Statt vier Testtexten wurden nun acht betrachtet, in denen die folgenden sieben Kategorien annotiert wurden: „Ort", „Relation", „Relationales Verb", „Raumbeschreibung", „Raumthema", „Raumhinweis" und „Raummetapher".

Jahrhundertspezifische Testreihe
Insgesamt schwanken die kategorienübergreifenden F1-Score-Werte in der jahrhundertspezifischen Testreihe zwischen rund 38 % (bei Stanišić mit dem Trainingskorpus des 18. Jahrhunderts) und rund 64 % (bei Josten mit dem Trainingskorpus des 21. Jahrhunderts) und damit um insgesamt rund 26 %. Diese Werte zeigen einerseits, dass mit dem sieben-Kategorien-Modell insgesamt höhere Quoten erreicht werden können; der höchste Wert dieser Testreihe liegt mit den bereits erwähnten 64 % um 6 % höher als der höchste Wert der ersten Testreihe für das fünf-Kategorien-Modell. Auf der anderen Seite ist dieses Modell aber auch anfälliger für Schwankungen, die mit dem literarhistorischen Kontext zusammenhängen. Das sieben-Kategorien-Modell scheint stärker an zeitliche Spezifika gebunden und damit weniger generisch einsetzbar zu sein (vgl. Abbildung 6.4).

Da eine Testbasis von vier Texten zu klein ist, um eine solche Hypothese zu stützen, wurden an dieser Stelle acht Testtexte einbezogen. Von diesen acht Texten, erreichen fünf die besten Ergebnisse mit Trainingsdaten des eigenen Jahrhunderts, zwei mit Trainingsdaten des angrenzenden Jahrhunderts, und wieder ist es nur der *Geisterseher* von Schiller, der die beste Quote mit einem Modell erreicht, das mit 40.000 annotierten Tokens aus dem 20. Jahrhundert trainiert wurde, also mit Daten aus dem übernächsten Jahrhundert.

Auch in dieser Testreihe (vgl. Abbildung 6.4) zeigen sich die Kategorien „Ort" und „Relation" als diejenigen, die für das Modell am besten erkennbar sind. Es folgen „Raumhinweise" und „relationale Verben", wobei mal die eine und mal die

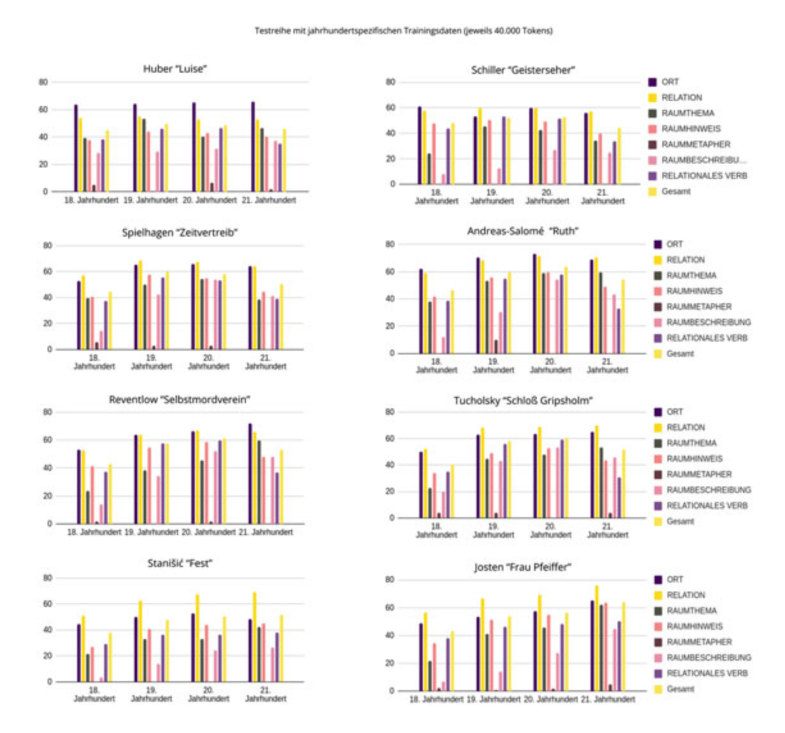


Abbildung 6.4 Visualisierung der Testergebnisse des CRF-Modells mit sieben Raumkategorien

andere Kategorie etwas besser erkannt wird. Ähnlich ist es bei den im Ranking dieser Auswertung an nächster Stelle stehenden Kategorien „Raumthemen“ und „Raumbeschreibungen“. In einem Testtext wird die eine, im nächsten die andere Kategorie besser identifiziert. Durchgehend schlecht werden Metaphern erkannt. Das könnte daran liegen, dass diese Kategorie bisher lediglich als Störfaktor mitannotiert und ausgewertet wurde. Hier wurde in der dritten Iteration die Annotationen nachgebessert.

Testreihe mit kumulierten Daten
Analog zur Testreihe der ersten Iteration wurden auch hier auf dieselbe Weise Durchschnittswerte der Testtexte ermittelt. Abweichend von der ersten Testreihe der ersten Iteration wurden an dieser Stelle acht Testtexte in die Berechnung einbezogen. Es ergibt sich also ein auf einer größeren Datenbasis fundiertes Bild.

Tatsächlich zeigt sich für das sieben-Kategorien-Modell ein ähnliches Bild wie für das fünf-Kategorien-Modell. Die Kurven steigen mit einigen Ausnahmen nahezu linear an (vgl. Abbildung 6.5). Allerdings steigen die Werte langsamer an als beim fünf-Kategorien-Modell, was dadurch erklärt werden kann, dass zum Teil bei mehr Kategorien die Anzahl der Beispiele im Trainingskorpus geringer ist (viele relationale Verben fielen z. B. vorher in die Kategorie Relation).

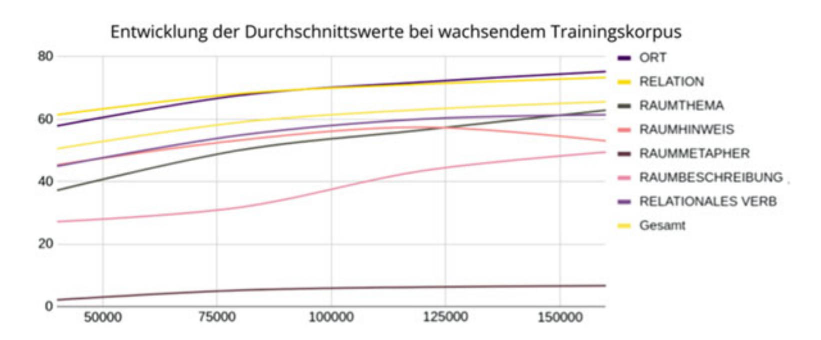


Abbildung 6.5 Durchschnittswerte von vier Testtexten mit wachsendem Trainingskorpus

Besonders interessant ist auch, dass der Durchschnittswert des Gesamt-F1-Wertes aller Kategorien ähnlich ist. Erreicht das fünf-Kategorien-Modell hier einen Wert von 61 %, so sind es beim sieben-Kategorien-Modell 65,56 %. Hier lohnt sich aber ein genauer Blick, denn die neu eingeführten Kategorien erreichen zwar selbst mit rund 49,43 % und 61,41 % F1-Score im Durchschnitt bei einem 160.000-Tokens-Trainingskorpus noch keine zufriedenstellenden Werte, die anderen Kategorien profitieren aber von der Ausdifferenzierung. Vier der fünf Kategorien des fünf-Kategorien-Modells zeigen eine Verbesserung. Nur die Erkennung der Metaphern sinkt insgesamt von 11 % im fünf-Kategorien-Modell auf 6,63 % im 7-Kategorien-Modell. Die Erkennungsgenauigkeit der Orte steigt dagegen von rund 74 % auf rund 75 %. Die Erkennung der Relationen verbessert sich um rund 10 Prozentpunkte auf 73,35 %, die der Raumthemen um rund 6 Prozentpunkte auf 62,86 % und die der Raumhinweise ebenfalls um rund 6 Prozentpunkte auf 53,07 %.

Resultate für das weitere Vorgehen

Die Testreihen nach der zweiten Iteration haben bestätigt, dass die Erweiterung des Raummodells auf sieben Kategorien sinnvoll ist, um die Konzepte zu schärfen. Das sieben-Kategorien-Modell wird darum beibehalten. Um die Erkennungsgenauig-

keit auf einen Zielwert von durchschnittlich 70 % Gesamterkennungsgenauigkeit zu bringen, muss das Trainingskorpus aber noch weiter ausgeweitet werden. In der dritten Iteration wurden darum weitere Trainingsdaten annotiert.

6.2.3 Die dritte Iteration (Ausweitung des Trainingskorpus)

In der dritten Iteration wurde das Trainingskorpus um insgesamt 80.000 Tokens auf 240.000 Tokens erweitert. Beim Aufbau des Trainingskorpus wurde dabei weiterhin auf Diachronie geachtet, d. h. die neuen Abschnitte wurden so eingefügt, dass die ersten 60.000 Tokens aus dem 18. Jahrhundert stammten, die nächsten 60.000 aus dem 19. Jahrhundert usw. Wie bei den Testreihen der ersten und zweiten Iteration wurden die Testtexte in jeder Erweiterungsphase getestet, sodass am Ende jeder Testtext erst mit einem Modell geprüft wurde, das mit 60.000 Tokens aus dem dem Veröffentlichungskontext entsprechenden Jahrhundert trainiert wurde, dann mit einem, das mit 120.000 Tokens aus dem 18. und 19. Jahrhundert trainiert wurde, dann mit einem auf 180.000 Tokens aus dem 18.–20. Jahrhundert trainierten und schließlich mit einem alle Jahrhunderte umfassenden Modell.

Besonders eklatant ist hier, dass die jahrhundertübergreifende Kumulierung der Trainingsdaten kaum eine Steigerung der Erkennungsgenauigkeit erreicht. Ein Blick auf die Small Multiples der Testreihe (Abbildung 6.6) zeigt, dass zwar die Erweiterung der jahrhundertspezifischen Trainingskorpora von 40.000 auf 60.000 Tokens zu einigen Prozentpunkten Steigerung führt, danach verlaufen die Kurven aber nahezu horizontal. Bei manchen Kategorien und / oder dem ein oder anderen Testtext folgt sogar eine Verschlechterung der Quote, wenn Trainingsdaten aus einem anderen Jahrhundert zum Trainingskorpus hinzugefügt werden (so z. B. bei Stanišić und Josten, bei denen allerdings auch nach dem Durchlauf mit dem 60.000er-Modell ein Test mit einem Modell durchgeführt wurde, das 120.000 Tokens aus dem 18. und 19. Jahrhundert umfasst, also der Zeitspanne, die am weitesten vom eigenen Kontext entfernt ist). Diese Daten stützen die bereits geäußerte Vermutung, dass es sich bei der Kategorie des literarischen Raumes zumindest teilweise um eine Zeitkontext-kritische Analysekategorie handelt.

In dieser Testreihe (vgl. Abbildung 6.6) erreichte bereits einer der Testtexte das Ziel von 70 % Gesamterkennungsgenauigkeit. Mit 73,5 % lag Andreas-Salomés *Ruth* sogar um 3,5 Prozentpunkte über dem Zielwert, was unter anderem auch daran gelegen haben könnte, dass dieser Text als letztes und damit in relativer zeitlicher Nähe zu den neuesten Trainingsdaten manuell annotiert wurde. Aus diesen Beobachtungen wurden für die nächsten Iterationen wieder zwei Schlussfolgerungen gezogen.

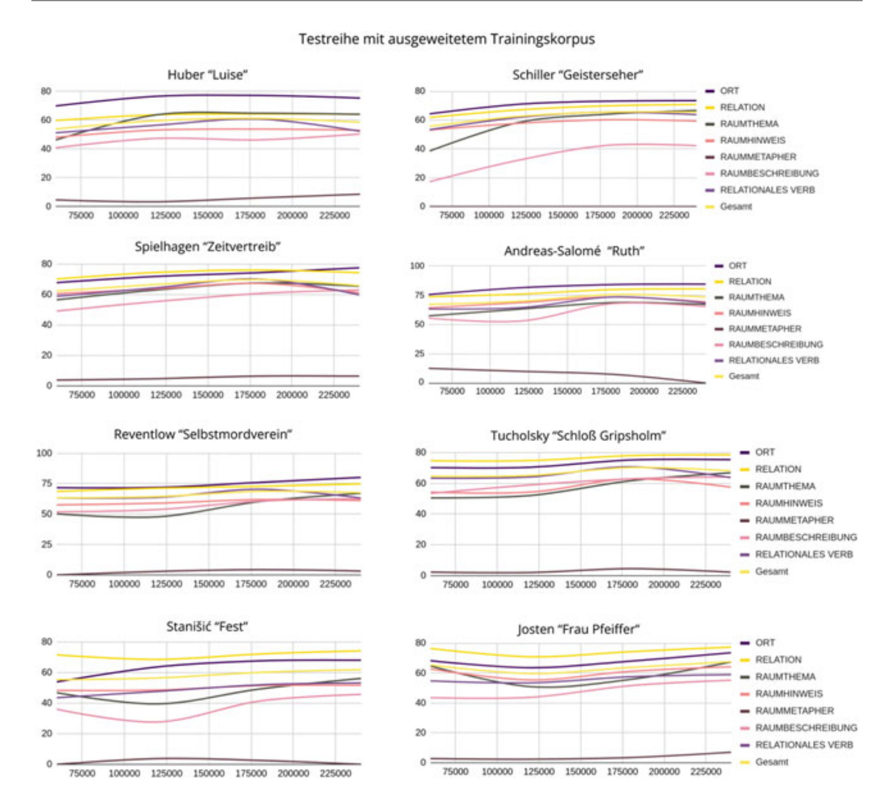


Abbildung 6.6 Small Multiples der Daten, die während des Ausweitungsprozesses des Trainingskorpus erhoben wurden

Resultate für das weitere Vorgehen
Zuerst einmal legen die vergleichsweise guten Daten des zuletzt annotierten Testtextes nahe, dass während des Annotationsprozesses unterschiedlich annotiert wurde. Die Erfahrung, dass die Kategorien während des Annotierens geschärft und dadurch aber auch verändert werden können, bekräftigt diese Vermutung. Das legt allerdings auch den Schluss nahe, dass das Trainingskorpus nach der dritten Iteration relativ heterogen und wahrscheinlich auch insgesamt inkonsistent annotiert worden ist. In der nächsten Iteration wurden darum alle Annotationen nochmals geprüft. Eine höhere Konsistenz wurde angestrebt, wozu auch gehört, dass die den Kategorien zugrundeliegenden Konzepte während der Prüfung weiter geschärft wurden. Diese Schärfung wurde in neuen, genaueren Guidelines festgelegt. Anhand dieser Guideli-

nes wurden die Testtexte erneut ausschnittsweise manuell annotiert. Die Ausschnitte umfassen jeweils 10.000 Tokens.

Ein zweiter Schluss aus der Auswertung dieser Testreihe ist, dass erneute Erweiterungen des Trainingskorpus jahrhundertspezifisch sein sollten. Modelle, die mit Daten des eigenen und höchstens noch des angrenzenden Jahrhunderts trainiert wurden, haben die höchste Wahrscheinlichkeit zu guten Ergebnissen zu führen. Vor einer erneuten Ausweitung standen aber erst einmal die Erhöhung der Konsistenz und die Entwicklung neuer Guidelines, die im folgenden Kapitel dokumentiert werden.

6.2.4 Die vierte Iteration (Konsistenzerhöhung und Guidelines)

In der vierten Iteration wurde zuerst am Trainingskorpus gearbeitet und dann ein Test an den unveränderten Testtexten durchgeführt. Da sich herausstellte, dass das nun konsistentere Trainingskorpus sich negativ auf die Erkennungsraten in den unveränderten Testtexten auswirkte, wurden die Regeln, die für die Konsistenzerhöhung entwickelt wurden, als präzisere Guidelines formuliert. Mit Hilfe dieser präziseren Guidelines wurden die Testtexte dann ausschnittsweise erneut annotiert. So konnten am Ende deutlich bessere Ergebnisse erzielt werden als in den Testreihen der ersten Iterationen.

Die Konsistenzerhöhung im Trainingskorpus wurde in zwei Schritten durchgeführt. Zunächst wurden alle Annotationen aller Kategorien geprüft und verbessert. Bei Ambiguitäten wurde der Versuch unternommen, nach genauen Regeln grundsätzliche Entscheidungen zu treffen, welche der Kategorien, von denen meist zwei in seltenen Fällen auch drei in Frage kommenden zur Annotation genutzt wurde. Dies betraf z. B. die grundsätzliche Entscheidung, dass Orte als geographisch feste Größen definiert wurden und nicht als Räume, in oder an denen Figuren sich aufhalten können. Eine andere Grundsatzentscheidung, die zu diesem Zeitpunkt getroffen wurde, ist, dass Raumbeschreibungen nur dann als solche annotiert werden, wenn es sich dabei nicht um das Gerundium eines Verbs (wie z. B. „gekrümmt“) handelt. Obwohl auch solche Verbformen beschreibend eingesetzt werden, wird die Kategorie „relationale Verben“ bevorzugt, da es sich dabei um die konkretere Raumkategorie handelt, die auch in der Berechnung des RIWs (vgl. Kapitel 7) einen höheren Wert zugewiesen bekommt als Raumbeschreibungen. Dies hat den Grund, dass auf relationale Verbformen zurückgehende Beschreibungen wie z. B. „gebogen“ oft klarer räumlich sind als Raumbeschreibungen, die auch häufig implizit auf Räume referenzieren, wie z. B. in der Phrase „seine exotische Ehefrau“. Die davon abstrahierende allgemeine Regel, dass stets die konkretere Raumkategorie

in Zweifelsfällen gewählt wird, hat eine Ausnahme: Ausdrücke, die zu „Raumthemen“ beitragen, gehören häufig auch zu einer konkreteren Raumkategorie. Hier wird aber stets die Kategorie „Raumthemen“ bevorzugt, da diese literaturwissenschaftlich besonders relevant ist, wie der Forschungsüberblick in Abschnitt 2.2.2 gezeigt hat.

Im zweiten Schritt wurden einzelne Wörter, die sich im ersten Schritt als besonders häufig mehrdeutig in Bezug auf die Raumkategorien erwiesen haben, im Trainingskorpus per Suchfunktion ausfindig gemacht und erneut geprüft. In diesem Schritt wurden nicht nur die neuen Regeln erneut auf konsistente Anwendung geprüft (wodurch an der Precision des Trainingskorpus gearbeitet wurde), es konnten auch Vorkommnisse ausfindig gemacht werden, die in den ersten Annotationsdurchgängen trotz mehrerer Iterationen übersehen wurden. Diese wurden also nachträglich annotiert, wodurch das Trainingskorpus auch im Hinblick auf den Recall verbessert werden konnte. In dieser Iteration wurden auch die letztgültigen Annotationsguidelines wie sie nun in Kapitel 4 nachzulesen sind, erstellt.

Die Kategorie der Raummetaphern erwies sich auch hier wieder als eine besondere. Im Gegensatz zu den anderen Kategorien können Raummetaphern nur selten an einem konkreten Wort festgemacht werden. Deutlich öfter handelt es sich um Passagen oder sogar ganze Satzteile, in seltenen Fällen ist es ein ganzer oder sind es sogar mehrere Sätze, die metaphorisch genutzt werden. Aus diesem Grund war das Vorgehen zur Konsistenzerhöhung der Annotationen dieser Kategorie im Trainingskorpus ein abweichendes. Zunächst wurden alle im Trainingskorpus annotierten Metaphern extrahiert. Es wurde festgestellt, ob es sich um eine ein-Wort-Metapher handelt, oder um eine Phrase oder ein Raumbild. Bei allen Metaphern wurde jeweils ein Wort ausgemacht, dass die raummetaphorische Bedeutung trägt, und als Kernwort festgehalten. Außerdem wurde die Bedeutung der jeweiligen Metaphern erfasst. Um all diese Komponenten später analysieren zu können, wurden sie in eine relationale Neo4J-Graphdatenbank eingetragen. Hierin ist es möglich, nicht nur die Komponenten der Metaphern festzuhalten, sondern die Metaphern auch noch mit den Titeln der Romane zu verknüpfen, in denen sie vorkommen. Auf diese Weise entstehen Relationen, die es erleichtern, die in einem bestimmten Zeitrahmen besonders häufig genutzten Metaphern ausfindig zu machen. Außerdem wurde ein zweiter Relationstyp hinzugefügt: Wenn eine Metapher die Variation einer anderen ist, wie z. B. bei den beiden Metaphern „in schlechte Hände geraten“ und „in schlechten Händen sein“, so wurde dafür eine Variationsrelation (benannt als „Variation“) angelegt. Um nun die Konsistenz der Metaphern-Annotation im Trainingskorpus zu erhöhen, wurden zwei Abfragen gemacht. Zuerst wurden besonders häufige Kernworte von Raummetaphern ausfindig gemacht. Dazu gehören z. B. Hände, Kopf und Herz. Dann wurden Metaphern ausfindig gemacht, die in besonders vielen,

im Trainingskorpus enthaltenen, Romanpassagen vorkommen. Dazu gehören z. B. groß, Höhe und fesseln. Insgesamt wurde eine kleine Stichprobe der sechs häufigsten Kernworte und der sechs häufigsten Metaphern notiert und das Trainingskorpus dann nach diesen untersucht. Auf diese Weise wurden zuvor übersehene raummetaphorische Ausdrücke annotiert. Im Idealfall wäre eine solche Überarbeitung für jede Metapher durchgeführt wurden. Bei einer Anzahl von insgesamt rund 800 Raummetaphern war dies im zeitlichen Rahmen einer Studie, in deren Fokus nicht-metaphorische Raumausdrücke stehen, leider nicht möglich. Die hier entstandene relationale Datenbank *laRa* mit raummetaphorischen Ausdrücken und ihren Relationen zu literarischen Texten und zueinander wird aber stetig weiter verbessert, für weitere Forschung zu einer digitalen Metaphorologie genutzt und zur Nachnutzung für Anschlussforschung vorgehalten und bereitgestellt.[4] Wenn auch die Verbesserung der automatischen Erkennung von Raummetaphern in dieser Arbeit nicht weiter verfolgt wird, so können die fortlaufend verbesserten Daten doch zu einem späteren Zeitpunkt wieder zu einem Machine-Learning-basierten Ansatz zurück geführt werden.

An dieser Stelle muss offengelegt werden, dass es als Einzelperson nicht möglich ist, ein absolut durchgängig konsistentes Trainingskorpus im Umfang des hier angefertigten zu erstellen. Darum betont z. B. auch Reiter (2020) in seinen Empfehlungen für Annotationsrichtlinien, dass immer mehr als eine Person annotieren sollte. Dennoch wurde das Forschungssetting dieser Studie gerade so angelegt, dass die Annotationen im Trainingskorpus als Einzelarbeit erstellt wurden. Dadurch ergibt sich einerseits eine größere methodische Nähe zur nicht-digitalen geisteswissenschaftlichen Forschung, die oft einzeln durchgeführt wird. Andererseits kann auf diese Weise herausgefunden werden, inwiefern die Methode des maschinellen Lernens auch in solchen Einzelforschungssettings angewendet werden kann. Das hier erstellte Trainingskorpus verstehe ich – im Rahmen der digitalen Literaturwissenschaften – als einen ersten Schritt, einen Auftakt für ein annotiertes Trainingskorpus für CRF-Algorithmen zur automatischen Erkennung von Räumen in literarischen Werken. Alle Trainings- und Testdaten wurden – soweit urheberrechtlich möglich – in einem GitHub-Repository frei zur Verfügung gestellt, sodass weiter daran gearbeitet werden kann. Obwohl aber das Trainingskorpus und die damit trainierten Modelle in dem Bewusstsein eingesetzt wurden, dass sie noch nicht ideal aufbereitet sind, konnten damit doch bereits beachtliche Ergebnisse erzielt werden. Ich deute dies als Hinweis darauf, dass die automatische Erkennung von Ausdrücken,

---

[4] Eine erste Anschlussstudie zur Raummetapher als Störfaktor der automatischen Erkennung von Indikatoren narrativer Raumdarstellung wurde bereits auf der DHd2022-Konferenz vorgestellt (Schumacher, 2022).

die zur Darstellung von Raum in literarischen Texten beitragen und die zur (Re-) Konstruktion der storyworld dienen (auch wenn weder die Konstruktion durch einen Schreibenden noch die mentale Rekonstruktion durch Lesende im Fokus dieser Arbeit stehen) umsetzbar und, in nicht allzu ferner Zukunft, lösbar ist.

Resultate der vierten Iteration und weiteres Vorgehen
Die zentrale Erkenntnis aus dieser Iteration war überraschend. Die Konsistenzerhöhung des Trainingskorpus führte zunächst zu einem Einbruch der Erkennungsraten. Nur bei zwei der insgesamt acht Testtexte führte die Konsistenzerhöhung im Trainingskorpus auch zu einer Verbesserung der Gesamterkennungsrate. Bei einem Testtext blieb diese nahezu unverändert. Bei fünf Testtexten verschlechterte sich die Gesamterkennung von −0,63 % (Tucholsky) bis hin zu −7,81 % (Stanišić). Die Überarbeitung der Testtexte nach den spezifizierten Guidelines brachte dann für fünf Texte wieder eine Verbesserung der Erkennungsraten. Die anderen drei Testtexte zeigten eine vergleichsweise leichte Verschlechterung um rund 1–2 Prozentpunkte. In diesen Daten deutet sich an, dass auch das stärker inkonsistente Trainingskorpus und die gleichsam inkonsistent annotierten Testtexte zusammen ein System bilden, dass so regelhaft funktioniert, dass das mit diesen Trainingsdaten trainierte Modell insgesamt eine ähnlich gute Erkennungsrate erreicht wie das mit konsistenteren Daten trainierte Modell. Dieser überraschende Effekt deutet darauf hin, dass auch intuitiver gesetzte Annotationen auf Gesetzmäßigkeiten beruhen, die ein Algorithmus erlernen kann, und dass auch dabei irgendwann eine kritische Masse erreicht wird, in der Regelmäßigkeiten / Muster ausgebildet werden, die der Algorithmus erkennt. Eine Beschreibung dieses Effekts liegt in der bisherigen Forschungsliteratur zur Annotation von Trainings- und Testdaten meines Wissens noch nicht vor, was daran liegen könnte, dass inkonsistente Trainings- und Testdaten meist von Beginn an vermieden werden z. B. indem man einen überindividuellen Goldstandard erstellt. Auch an dieser Stelle kann dieser Effekt nicht weiter untersucht werden, da das zu weit vom Kerninteresse dieser Studie weg führen würde. Insgesamt ergab die Konsistenzerhöhung von Trainingskorpus und Testtexten eine Verbesserung der durchschnittlichen Gesamterkennungsgenauigkeit von 1,2 Prozentpunkten. Damit liegt die durchschnittliche Gesamterkennungsgenauigkeit nach dieser Iteration bei 65,72 %. Um die restlichen rund 5 Prozentpunkte bis zur Zielmarke von 70 % zu erreichen, wurden weitere Trainingsdaten erstellt. Diese wurden mit dem bisher besten Modell mit der durchschnittlichen Erkennungsgenauigkeit von rund 65 % vorannotiert, dann manuell verbessert und zum Trainingskorpus hinzugefügt.

Ein hier bisher nicht erwähnter, da erfolgloser und darum verworfener Schritt, bestand darin, die Trainingsdaten auszuweiten, indem automatisch mit dem bisher besten CRF-Modell weiteres Trainingsmaterial annotiert wurde. Die daraufhin

durchgeführten Tests zeigten eine Verschlechterung durch das mangelhafte, da nur zu rund 65 % mit einer manuellen Annotation übereinstimmende und damit auch zum restlichen Trainingskorpus passende, neue Trainingsmaterial. Während dieses erfolglosen Zwischenschrittes passierte anfangs allerdings ein Fehler, der hier kurz dokumentiert werden soll, da die (hier aus Versehen ausgewählte) Vorgehensweise zu Verbesserungen von Modellen führen kann, die auf kleinen Trainingskorpora trainiert werden. Bei der Erstellung des neuen Trainingsmaterials (80.000 automatisch annotierte Tokens, 20.000 aus jedem Jahrhundert) wurde hier nämlich zunächst statt des neuen automatisch annotierten Trainingsmaterials für das 20. Jahrhundert das bisher genutzte, manuell erstellte Trainingsmaterial zum zweiten Mal zum Trainingskorpus hinzugefügt. Die Erkennungsrate des mit dem neuen Material (60.000 automatisch annotierte Tokens, 20.000 duplizierte manuell annotierte Tokens) trainierten Tools stieg um rund 2 Prozentpunkte auf 67,82 %. Es wurden weitere 80.000 Tokens automatisch annotiertes Trainingsmaterials hinzugefügt (dieses Mal passierte kein Fehler) und die durchschnittliche Erkennungsgenauigkeit fiel wieder auf 56,8 % zurück. Nach weiteren Tests konnte festgestellt werden, dass das automatisch annotierte Trainingsmaterial eher eine Verschlechterung der durchschnittlichen Erkennungsgenauigkeit bewirkte und dass die anfängliche Verbesserung auf das doppelte, manuell annotierte Trainingsmaterial aus dem Trainingskorpus des 20. Jahrhunderts zurückgeführt werden konnte. Darum wurden weitere Tests mit verdoppeltem Trainingsmaterial durchgeführt. Durch Verdopplung des Trainingskorpus von 240.000 auf 480.000 Tokens verbesserte sich die Erkennungsgenauigkeit bei den acht Testtexten um 0,65–5,5 Prozentpunkte, die Steigerung fiel also recht unterschiedlich stark aus. Die Verbesserung durch Hinzufügen weiterer 80.000 manuell annotierter Tokens fiel deutlich höher aus. Außerdem verlor sich der Effekt der Verbesserung durch verdoppeltes Trainingsmaterial bereits bei 320.000 Tokens wieder – verdoppelt man hier auf 640.000 bleiben die Ergebnisse nahezu gleich.

Nun lässt sich an dieser Stelle noch die oben stehende Beobachtung, dass inkonsistente Trainingsdaten und inkonsistente Testdaten zusammen ähnlich gut wirken wie konsistentere Trainingsdaten mit konsistenteren Testdaten, präzisieren. Da die menschlich inkonsistente Annotation eher intuitiv anders interpretierte Kategorien betrifft (die dann beim Test als false positives gewertet werden könnten), kann sie insgesamt ein regelmäßiges System bilden. Das CRF-Modell, das mit einer bestimmten Erkennungsgenauigkeit annotiert, produziert dabei hingegen vergleichsweise viele Fehler („echte" false positives und vor allem sehr viele false negatives). Statt z. B. in einer „Treppe" einen Raumhinweis zu sehen anstatt eines Wortes, dass zum Thema der Öffnung und Schließung beiträgt, könnte die Software einfach ein false negative produzieren, indem „Treppe" gar nicht als einer Raumkategorie zugehörig annotiert wird. Menschliche Annotator*innen – so hat auch die Auswertung des Inter

Annotator Agreements (dokumentiert in Abschnitt 6.2) ergeben – entscheiden zwar oft unterschiedlich, machen aber vergleichsweise wenig echte Fehler. Darum wurde von der Ausweitung des Trainingsmaterials mit Hilfe von automatisch annotierten Trainingsdaten auch wieder Abstand genommen. Am Ende sind manuell erstellte Trainingsdaten die bessere Wahl.

6.2.5 Die fünfte Iteration (erneute Ausweitung des Trainingskorpus)

Um nun die letzten Prozentpunkte zur 70 %-Hürde noch zu schaffen, wurden weitere Trainingsdaten manuell nach den spezifizierten Guidelines annotiert. Diese neuen Trainingsdaten umfassen 80.000 Tokens, die zu gleichen Teilen Romanen aus dem 18., 19., 20. und 21. Jahrhundert entnommen wurden. Dieses manuell annotierte Trainingsmaterial wurde in das Trainingskorpus integriert, sodass dieses am Ende 320.000 manuell annotierte Tokens umfasste. Dadurch konnte die durchschnittliche Gesamterkennungsgenauigkeit schließlich auf 73,99 % gebracht werden.

Ganz zu Anfang des CRF-Trainingsprozesses zeigten Testergebnisse, dass die Integration der Kategorie „Metapher“ eine Verbesserung der Erkennung der anderen Kategorien mit sich brachte. Der Störfaktor der uneigentlichen Raumausdrücke schien sich dadurch klein halten zu lassen. Nachdem während des Trainingsprozesses aber festgestellt wurde, dass sich die Erkennung für die Kategorie „Metapher“ sowohl durch Konsistenzerhöhung des Trainingskorpus als auch durch Hinzufügen weiteren Trainingsmaterials kaum verbessern ließ, wurden erneut Tests durchgeführt, um diesen Faktor zu überprüfen. Hier zeigte sich nun, dass die Berücksichtigung der Kategorie „Raummetaphern“, die am Ende durchschnittlich auch nur 7,74 % Erkennungsgenauigkeit aufweisen konnte, innerhalb des finalen, mit 320.000 Tokens trainierten Modells keine Verbesserung, sondern, im Gegenteil, eine Verschlechterung der Erkennung der anderen Kategorien bewirkte. Die durchschnittliche Gesamterkennungsgenauigkeit des Modells betrug ohne die Metaphern-Kategorie am Ende 75,65 % und damit 1,66 Prozentpunkte mehr als beim Modell, in dem Metaphern integriert waren.[5] Beide Raum-Classifier – der die Metaphern-

[5] Es gibt zwei Möglichkeiten, die Metaphern-Kategorie aus dem Trainingskorpus zu eliminieren, die hier beide getestet wurden. Werden die Annotationen nur herausgelöscht und durch die Nullkategorie ersetzt, so verbessert sich die Erkennung der anderen Kategorien bereits. Werden die bisher als Metaphern annotierten Ausdrücke aber in die anderen Kategorien klassifiziert als wären sie im eigentlichen Wortsinne verwendet worden, so fällt diese Verbesserung deutlich höher aus. In den hier dokumentierten Tests verbesserte sich die durchschnittliche Erkennungsgenauigkeit durch Löschen der Metaphern-Annotationen von 73,99 %

Kategorie enthaltende und der lediglich sechs Raumkategorien umfassende – werden zur Nachnutzung frei zur Verfügung gestellt und können über ein GitHub-Repository oder über Zenodo heruntergeladen werden (Schumacher, 2021c).

Insgesamt fallen im Trainingskorpus 7,7 % der Annotationen in die Kategorie der Raummetapher. Das sind 4.754 von 320.000 Tokens. Runtergerechnet auf die Länge eines eher kurzen Romans von 50.000 Tokens sind das immerhin rund 743 Wörter. Hinzu kommen jahrhundertspezifische Schwankungen der Häufigkeit in der Verwendung von Raummetaphern. Diese wurden bereits bei den ersten Tests des Modells mit fünf Kategorien erfasst, sodass hier festgehalten werden kann, dass in den Romanen des 18. Jahrhunderts die meisten und in denen des 20. Jahrhunderts am seltensten Raummetaphern verwendet wurden (vgl. Abbildung 6.7).[6]



Abbildung 6.7 Annotationskategorien im Trainingskorpus pro Jahrhundert

Raummetaphern sind also vor allem bei der Betrachtung literarischer Texte des 18. und 19. Jahrhunderts eine nicht zu vernachlässigende Größe. Obwohl sie (derzeit noch) nicht mit Hilfe eines Machine-Learning-Tools erfasst werden können, sollten

auf 74,77 %. Die Umkategorisierung ergab eine Steigerung auf 75,64 %. Die Erhöhung fiel also rund doppelt so hoch aus (1,66 Prozentpunkte mehr statt 0,78).

[6] Eine Übersicht über die Anzahl der Annotationen im Trainingskorpus insgesamt und in den jahrhundertspezifischen Abschnitten ist im GitHub-Repository zum Raum-Classifier einsehbar (Schumacher, 2021c).

sie also trotzdem Berücksichtigung finden. Hier helfen Mixed-Methods-Ansätze im Allgemeinen und die innerhalb dieser Studie erstellte Datenbank der Raummetaphern *laRa* im Besonderen. Auf welche Weise in der Analyse Raummetaphern und ihre Bedeutung für bestimmte Wörter, die auch zur Darstellung des tatsächlichen Raumes der storyworld genutzt werden, in die Analyse einbezogen werden können, zeigen die beispielhaften Fallstudien in Kapitel 8. Wenn die gesamten Raummetaphern im Trainingskorpus auch nur 7,7 % des Wortmaterials ausmachen, so kann der Anteil metaphorischer Verwendung bei manchen Einzelwörtern sehr viel höher ausfallen. Die Fallstudien in Abschnitt 8.1.4 zeigen, dass die Beispielwörter „Welt" und „Himmel" in den einzelnen Teilkorpora zu bis zu 35 % in metaphorischen Kontexten stehen.

Das finale, in der weiteren Untersuchung verwendete CRF-Modell ist also eines, das sechs Unterkategorien aufweist (Ort, Relation, relationales Verb, Raumbeschreibung, Raumhinweis, Raumthema) und auf einem komplett manuell annotierten und iterativ verbesserten Trainingskorpus von 320.000 Tokens Umfang basiert. Die Tokens wurden zu gleichen Teilen aus insgesamt 80 Romanen aus den Jahrhunderten 18–21 extrahiert.

6.3 Bewertung

Die ausführlichen, nach jeder Iteration durchgeführten Tests haben gezeigt, dass es möglich ist, ein komplexes, auf literaturwissenschaftlichen Theorien aufbauendes Raummodell so zu operationalisieren, dass eine automatische Annotation mit einer Gesamterkennungsgenauigkeit von durchschnittlich rund 75 % möglich ist. Die textspezifische Gesamterkennungsgenauigkeit reicht dabei von rund 66 % bis 82 %, die Standardabweichung beträgt 4,57. Der untere Extremwert wird bei Therese Hubers *Luise* erreicht, der obere bei Andreas-Salomés *Ruth*. Bei beiden Extremwerten handelt es sich um Ausreißer, die über bzw. unter dem um die Standardabweichung erhöhten bzw. reduzierten Mittelwert liegen. Die anderen Texte erreichen zwischen 72,89 % und 79,44 %. Wahrscheinlich ist, dass das Modell auf andere Texte angewendet meistens zwischen 71,08 % und 80,22 % (Mittelwert + / − Standardabweichung) erreicht, wobei es Ausreißer nach oben und nach unten geben kann.

Die einzelnen Kategorien erreichen zwischen rund 7 % und 80 % durchschnittliche Erkennungsgenauigkeit. Dabei werden die 7 % für die Kategorie der Raummetapher erreicht, die anfänglich ins Modell aufgenommen wurde, um nicht als Störfaktor auf die anderen Kategorien zu wirken, später dann aber wieder herausgenommen wurde. Die für diese Arbeit zentralen Kategorien erreichen zwischen rund

60 % und 80 % Erkennungsgenauigkeit. Die Rangfolge entspricht in etwa dem, was angesichts der Ex- bzw. Implizitheit der Raumkategorie zu erwarten war. Eine schlechtere Erkennung der Kategorie „Raumbeschreibung“ kann aber auch mit den vergleichsweise niedrigen Annotationszahlen im Trainingskorpus zusammenhängen, für „Raumhinweise“ gibt es in den Trainingsdaten allerdings ähnlich viele Beispiele wie für die besser erkannten Kategorien. Neben Implizitheit und Anzahl der Annotationen im Trainingskorpus könnte Heterogenität der Kategorie sich ebenfalls auf die Erkennungsgenauigkeit ausgewirkt haben. Am besten werden mit rund 80 % „Ort“ und mit 79 % „Relationen“ erkannt, was darauf hindeutet, dass dies die konkretesten Raumkategorien sind. Es folgen „Raumthemen“ mit rund 76 %. „Relationale Verben“ sind mit rund 75 % für das Modell schon weniger genau erkennbar. Die eher impliziten Raumkategorien „Raumhinweis“ und „Raumbeschreibung“ erreichen rund 72 % und 60 % Erkennungsgenauigkeit. Ein Vergleich aller Kategorien mit anderen Tools zur automatischen Erkennung ist, wie in Abschnitt 5.2 dargelegt, nicht möglich, da es keine anderen Tools gibt, die diese Kategorien aufgreifen. Ein Vergleich der Kategorie „Ort“ mit dem, was das NER-Modell von Faruqui und Padó (2010) bei der Entitätenerkennung in Sachtexten erreicht, zeigt aber, dass die hier erreichten durchschnittlichen 80 % Erkennungsgenauigkeit bereits ein guter F1-Wert sind, der mit Ergebnissen von State-of-the-Art-Tools mithalten kann (vgl. Faruqui and Padó, 2010; Hänig et al., 2014, 113 und Abschnitt 5.2 dieser Arbeit).

Um auch für die üblicherweise nicht in NER-Tools enthaltenen Kategorien zumindest vage abschätzen zu können, wie klar die Kategorisierung dieses Modells ist und wie der Computer im Vergleich zu menschlichen Annotator*innen damit umgeht, wurde im Rahmen des forTEXT-Projektes eine manuelle Annotation einer Textpassage von jeweils zwei voneinander unabhängigen Annotatorinnen durchgeführt. Aus diesen Daten wurde das Inter-Annotator-Agreement errechnet. Bei den Annotatorinnen handelt es sich um literaturwissenschaftlich sehr gut vorgebildete Studentinnen, die bereits Erfahrungen mit der digitalen Annotation haben. Grundlage der Annotation waren die Guidelines (vgl. Kapitel 4) zu den Kategorien des hier entwickelten Modells. Die beiden Studierenden annotierten zunächst eine kurze Textpassage von 4.000 Tokens und konnten danach Fragen zu den Guidelines stellen. Daraufhin wurden diese noch einmal nachgeschärft. Die Textgrundlage für das Inter-Annotator-Agreement umfasst vier Ausschnitte aus vier der in dieser Untersuchung verwendeten Test-Texte. Die Passagen umfassen jeweils 5.000 Tokens, der zu annotierende Text also insgesamt 20.000 Tokens, und decken alle vier hier betrachteten Jahrhunderte ab.

Nun gibt es derzeit in den digitalen Literaturwissenschaften noch kein allgemein gängiges Maß für das Inter-Annotator-Agreement, auch wenn bereits einige

im Bereich der Computerlinguistik getestet wurden (vgl. Reiter, 2020, 196; Artstein and Poesio, 2008). Hier wurde zur Berechnung Cohen’s Kappa verwendet, ein Maß, dass es zulässt, Annotationen von zwei Annotierenden und in zwei oder mehr Kategorien einzubeziehen (vgl. Artstein, 2017; Hammann et al., 2014). Dass es sich dabei nicht um eine ideale Berechnungsweise handelt, hat auch der hier durchgeführte Test belegt. Das Maß Cohens Kappa lässt es grundsätzlich zu, ein Inter-Annotator-Agreement auf Basis aller qualitativ belegter Kategorien (also hier die sieben Raumkategorien) zu berechnen. Es ist aber auch möglich, alles, was nicht in eine der Kategorien fällt, mit einzubeziehen (ich nenne die dadurch entstehende achte Kategorie hier Nullkategorie). Je nach Berechnung ist der Zufallsfaktor relativ gering oder sehr stark ausgeprägt. Wenn, wie im vorliegenden Fall, 18.341 von 20.000 Tokens in die Nullkategorie fallen, so ist die Wahrscheinlichkeit von Zufallstreffern hier sehr hoch. Um trotz der Problematiken dieser Berechnungen zu einem zielführenden Vorgehen zu kommen, wurde das Inter-Annotator-Agreement sowohl mit als auch ohne Nullkategorie berechnet. Anschließend wurde ein qualitativer Vergleich durchgeführt, um zu schauen, bei welchen Kategorien die Annotatorinnen eher dazu neigen, sie unterschiedlich zu annotieren. Dadurch wird aufgezeigt, bei welchen Kategorien Ambiguitäten wahrscheinlich sind, ein Phänomen, das von Berechnungen des Inter-Annotator-Agreements weitgehend unberücksichtigt bleibt.

Das Inter-Annotator-Agreement auf Basis der sieben Raumkategorien (ohne Nullkategorie) beträgt rund 0,71. Das beobachtete Agreement (observed Agreement) liegt bei 0,77 und die Zufallswahrscheinlichkeit bei 0,20. Das Problem bei dieser Form der Berechnung ist aber, dass die Fälle, in denen eine Annotatorin ein Wort als einer Kategorie zugehörig annotiert hat, während die andere der Meinung ist, dass es sich hier nicht um ein Wort handelt, das in eine Raumkategorie fällt, unberücksichtigt bleiben. Berechnet man Cohens Kappa nun mit Nullkategorie, so ergibt sich ein IAA von 0,43. Das Observed Agreement ist hier zwar mit 0,90 recht hoch, der Zufallsfaktor mit 0,84 aber ebenfalls.

Aus diesen Berechnungen lässt sich zunächst nicht viel mehr ableiten, als dass das Einverständnis der voneinander unabhängig Annotierenden weniger stark ist als der F1-Score, den das CRF-Modell erreicht (rund 0,75), wenn die automatischen mit den von mir händisch erstellten Annotationen abgeglichen werden. Wenn die Kategorien aber so mehrdeutig sind, dass die Übereinstimmung von zwei voneinander unabhängigen Literaturwissenschaftler*innen höchstens 0,71 beträgt, so kann es als vertretbar eingestuft werden, ein Tool zu nutzen, dass rund 0,75 Übereinstimmung mit den Annotationen einer Literaturwissenschaftlerin erreicht.

Beim qualitativen Vergleich der beiden Annotationen zeigt sich, dass die Kategorien „Ort“, „Raumhinweis“ und „Raumthema“ besonders nah beieinander liegen.

Dass es beim Raumthema öfter Unklarheiten gibt, scheint leicht erklärbar zu sein, denn die meisten Wörter, die in diese Kategorie fallen, sind grundsätzlich ambivalent. Es handelt sich z. B. häufig um relationale Verben, die aber typisch für ein Raumthema sind und darum zu dieser Kategorie gezählt werden, wie z. B. der Fall beim Verb „reisen“, das zum „Raumthema“ der Reise gehört. Auch Gegenstände, die einen Hinweis auf den Raum geben, wie z. B. „Wände“, werden zur Kategorie „Raumthema“ gezählt, da sie zur Beschreibung von Grenzen, Öffnung und Schließung beitragen. Die Kategorie „Raumthema“ ist in diesen Fällen also die übergeordnete. Bei der Kategorie „Ort“ überraschen die vergleichsweise häufigen Uneinigkeiten mehr, da es sich um eine der konkretesten Raumkategorien handelt. Der erste Eindruck wird allerdings dadurch relativiert, dass es hier nicht nur vergleichsweise häufig Uneinigkeit, sondern auch eine sehr hohe Übereinstimmung gibt. Außerdem gibt es relativ wenige Überschneidungen mit der Nullkategorie. Dennoch soll hier festgehalten werden, dass es vergleichsweise häufig vorkommt, dass eine Annotatorin ein Wort als „Ort“ klassifiziert, während die andere es als „Raumhinweis“ interpretiert. Besonders interessant sind auch die Ambivalenzen mit der Nullkategorie. Hier interpretiert eine der Annotatorinnen vor allem die Lexeme, die die andere als „Relation“ oder „relationales Verb“ annotiert hat, als Wörter, die zur Nullkategorie gehören. Die andere klassifiziert vor allem Wörter, die die Erste als „Raumhinweis“ oder „Raumbeschreibung“ (also die implizitesten Kategorien) annotiert hat, zur Nullkategorie. Hier zeigt sich wahrscheinlich eine Abweichung des jeweiligen (unterschwelligen) Raumverständnisses. Während eine der Annotatorinnen ein eher relationales Raumverständnis hat und darum häufiger relationale Ausdrücke annotiert als die andere, scheint diese eher zu einem absoluten Raumverständnis zu neigen und mehr Objekte (Raumhinweise) und deren Beschreibungen zu annotieren. Insgesamt sind es vor allem die relationalen und die impliziten Raumkategorien, die am stärksten variieren, wenn es um die Interpretation hinsichtlich der Zugehörigkeit zu Raumdarstellungen geht. Wenn etwas als Raumkategorie annotiert wird, herrscht aber eine zumeist recht große Übereinstimmung darin, in welche Kategorie ein Ausdruck fällt. Ausnahme ist hier die ambivalent angelegte Kategorie der „Raumthemen“. Auch die Kategorien „Ort“ und „Raumhinweise“ sind nicht immer so klar voneinander unterscheidbar wie die anderen Kategorien.

Ein Machine-Learning-Prozess wie der hier durchgeführte, kann selten als abgeschlossen betrachtet werden. Ein hauptsächlich von einer einzelnen Forscherin oder einem einzelnen Forscher durchgeführter Machine-Learning-Prozess hat darüber hinaus von vornherein immer noch großes Optimierungspotential, wenn es darum geht, das trainierte Tool für andere Projektsettings nutzen zu können. Die in dieser Arbeit erstellten Daten des Machine-Learning-Trainings werden darum – sofern rechtlich möglich – für Anschluss- und Nachnutzungsforschung zur Verfügung

gestellt und können so weiterhin fruchtbar gemacht werden, um die Annotation von Raumausdrücken in literarischen Texten nach dem hier entwickelten Modell zu erleichtern bzw. zu beschleunigen. Die öffentlich zugänglichen Ausschnitte des diachron aufgebauten Trainingskorpus können komplett genutzt oder auch jahrhundertgenau aufgespalten werden, um an Modellen zu arbeiten, die für spezifischere, nicht-diachrone Korpora besser funktionieren. Das hier entwickelte Modell und das damit trainierte Tool können aber auch eingesetzt werden, um einzelne der hier vorgestellten Kategorien in literarischen Texten automatisch annotieren zu lassen. Dafür eignet sich die Kategorie „Ort“ besonders, da die Erkennung bereits auf einem ähnlichen Niveau ist wie das der State-of-the-Art-Modelle aus dem Bereich der computerlinguistischen Entitätenerkennung.

7 Der Raumindexwert (RIW) und weitere Indexwerte

Ausgehend von den sechs Kategorien der Darstellung von Raum in Erzähltexten – Ort, Relation, relationales Verb, Raumhinweis, Raumbeschreibung und Raumthema – habe ich eine einfache Formel entwickelt, mit Hilfe derer für jeden Erzähltext im Korpus ein Raumindexwert (kurz: RIW) errechnet werden kann. Basis der Berechnung sind zwei Grundannahmen:

1. Wörter, die Rauminformationen vermitteln, machen einen Anteil des gesamten Wortmaterials von Erzähltexten aus.
2. Explizite Raumausdrücke referenzieren Rauminformationen stärker als implizite, da sie weniger von der Interpretation der Lesenden abhängen. Explizit und implizit werden hier nicht als rein binäre Werte verstanden, sondern als zwei Enden einer Skala, die eine Reihe von Zwischenwerten enthält.[1]

Aus der Grundannahme eins folgt, dass der prozentuale Anteil der Wörter und Phrasen, die eine raumreferenzierende Funktion haben, an der Gesamtwortmenge eines Erzähltextes ein Indikator für die Intensität ist, mit der von dem konkret vorliegendem Erzähltext Rauminformationen vermittelt werden. Die relative Berechnung ist hier ein wichtiger Faktor, um mit einzubeziehen, dass die Texte im Korpus unter-

[1] Nicht einbezogen werden hier implizite Verweise auf den Raum, die in keiner Weise an einzelnen Wörtern festgemacht werden können, sondern beim Lesen aus dem Satzzusammenhang erschlossen werden (vgl. Viehhauser, 2020). Das bedeutet zwar, dass der RIW nicht alle Facetten literarischer Raumdarstellung erfasst, da der Wert in dieser Analyse aber rein relational betrachtet wird, die RIWs der Texte also miteinander in Bezug gesetzt werden, können trotzdem quantitative Tendenzen der Raumdarstellung ausgemacht werden. Für den RIW bedeutet das aber, dass er als Basis- oder Mindestwert der Raumdarstellung interpretiert werden sollte. Rauminformationen, die aus dem Zusammenhang erschlossen werden, können hinzu kommen.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_7

schiedlich lang sind. Aus der Grundannahme zwei folgt, dass die Raumausdrücke nicht einfach quantitativ in Form natürlicher, gleichwertiger Zahlen summiert werden können. Darum werden sie anhand einer Skala von 0–1 gewichtet. Raumhinweise sind implizit und ihre Deutung in Bezug auf die Raumdarstellung hängt stark von der lesenden Person ab. Außerdem vermitteln sie neben Informationen über den Raum auch noch mindestens eine weitere Bedeutung, die sogar die primäre Bedeutung des jeweiligen Ausdrucks sein kann. Ihnen wird darum ein Wert von 0,5 zugeschrieben. Bei beschreibenden Raumausdrücken stehen Rauminformationen etwas stärker im Vordergrund als bei Raumhinweisen. Allerdings gibt es auch hier häufig sehr starke andere Bedeutungen. Der Satz „er fuhr ein großes Auto“ sagt zwar etwas über die räumliche Ausdehnung des Autos aus, kann aber z. B. auch etwas über den Status des Fahrers verraten. Beschreibenden Raumausdrücken wird hier darum der Wert 0,6 zugewiesen. Raumthemen sind expliziter als Raumhinweise und beschreibende Raumausdrücke, aber weniger konkret. Sie tragen mehr zur Raumdarstellung bei und Ausdrücke und Phrasen, die Raumthemen kommunizieren, dienen meist primär zur Vermittlung räumlicher Aspekte (auch wenn sie noch weitere Bedeutungsebenen haben können). Ihnen gebe ich den Wert 0,7. Relationen enthaltende Verben tragen viel zur Ausgestaltung des relationalen Raumes einer Erzählung bei. Aber auch hier schwingen oft andere Bedeutungsebenen mit. Das Verb „fortgehen“ z. B. drückt zwar eine relationale Bewegung aus, kann aber auch Aspekte wie Melancholie, Zwang oder auch Sehnsucht nach etwas Neuem beinhalten. Da die räumliche Komponente hier aber vergleichsweise stark und andere Bedeutungsebenen impliziter sind, bekommt diese Kategorie hier den Wert 0,8 zugewiesen. Sowohl Relationen als auch Ortsbezeichnungen können als explizite Raumdarstellungen gelten. Ob der Raum eher relational oder eher objektorientiert charakterisiert wird, ist, wie die Ausführungen in Abschnitt 2.3 gezeigt haben, von der theoretischen Perspektive abhängig. Da beide Seiten gleichwertige Argumente für einen relativen oder absoluten Raum vorgelegt haben, bewerte ich Relationen und Nennungen von Orten hier gleich. Beiden Kategorien wird also der Wert 1 zugewiesen. Die Einschätzung der Explizitheit und des daraus resultierenden Faktors folgt den Überlegungen zu den Raumkategorien in Abschnitt 2.1. Abbildung 7.1 zeigt die Bewertung im Überblick und ruft die visuellen Metaphern der Ortskategorien in Erinnerung.

Eine Quantifizierung der Raumeinheiten in Erzähltexten wird mit dieser Berechnung nicht angestrebt. Darum bleibt der Einwand von Genette, dass in detaillierten Beschreibungen (Genette und Knop, 1998, 118) Häufungen von Raumausdrücken auftreten können, die aber nur auf einen einzigen Ort verweisen, unberücksichtigt.

Der RIW jedes Textes im Korpus wird auf dieser Basis wie folgt berechnet[2]:

$$RIW = \frac{0,5 \times RH + 0,6 \times RB + 0,7 \times RT + 0,8 \times rV + R + O}{100 \times Wortzahl\ des\ Textes} \tag{7.1}$$

Raumkategorie	visuelle Metapher	Bewertung der Explizitheit
Ort		1
Relation		1
relationales Verb		0,8
Raumthema		0,7
Raumbeschreibung		0,6
Raumhinweis		0,5

Abbildung 7.1 Raumkategorien, visuelle Metaphern und Einschätzung des Explizitheitswerts

Mit Hilfe dieser Formel wird in der vorliegenden Studie jedem Text ein RIW zugewiesen, der theoretisch zwischen 0 und 100 liegen kann (auch wenn es im höchsten Maße unwahrscheinlich ist, dass die Extremwerte erreicht werden). Anhand ihrer Raumwerte kann die Ausprägung der räumlichen Dimensionen in Texten miteinander verglichen werden. Der Vergleich der Raumwerte kann einen Hinweis darauf geben, ob in einer bestimmten literarhistorischen Zeitspanne der relative Anteil an Raumdarstellung tendenziell größer war als in anderen. Außerdem können die RIWs anzeigen, ob es eine generelle Tendenz des Anteils von Raumausdrücken in Romanen gibt oder ob die räumliche Dimension individuell sehr unterschiedlich stark ausgeprägt ist.

Zwei Probleme treten bei der Anwendung dieser Formel auf, wenn die Datenbasis, wie hier, durch die eingesetzte Methode der automatisierten Erkennung nicht

---

[2] mit RH=Anzahl der Raumhinweise, RB=Anzahl der Raumbeschreibungen, RT=Anzahl der zu Raumthemen gehörigen Ausdrücke, rV=Anzahl der relationalen Verben, R=Anzahl der Relationen und O=Anzahl der Orte

akkurat ist. Erstens werden nicht alle Kategorien gleich gut erkannt. Zweitens funktioniert die Erkennung nicht bei allen Texten gleich gut. Die erste Problematik ist eine der Ambivalenzen der Raumkategorien, die zweite eine der Standardisierung der Referenzierung von Rauminformationen im Roman. Die Ambivalenzproblematik, so hat sich durch den Abgleich von F1-Score und Inter-Annotator-Agreement gezeigt, liegt hier in einem Umfang vor, der dem der (Un-)Einigkeit zwischen zwei Literaturwissenschaftler*innen entspricht. Einfach ausgedrückt, ist das Ergebnis der automatischen Annotation, gemessen an der Treffergenauigkeit, ebenso gut wie die Annotation eines anderen menschlichen Annotators oder einer Annotatorin. Dieser Faktor wird hier darum nicht weiter beachtet. Stattdessen werden die erzielten Annotationen als zu 100% genau (für einen abweichenden Annotierenden) angenommen. Um den zweiten Störfaktor besser einschätzen zu können, wurde der RIW zweier Testtexte berechnet. Einer dieser Testtexte, Schillers *Geisterseher* erreichte die höchste Übereinstimmung mit der automatischen Annotation, der andere, Hubers *Luise* die niedrigste. Der RIW bei manueller Annotation von Schillers *Geisterseher* liegt bei 10,34, bei automatischer Annotation bei 10,85. Die Abweichung ist erwartungsgemäß sehr niedrig. Der RIW bei manueller Annotation von Hubers *Luise* liegt bei 5,63, bei automatischer Annotation bei 9,63. Das heißt, es ist mit einer Abweichung der RIWs um in der Regel bis zu 4 Prozentpunkte zu rechnen. Diese Ungenauigkeit wird hier in Kauf genommen, um das Bild einer generellen Tendenz zeichnen zu können. Im Hinblick auf Nachnutzung des Tools wurde hier außerdem ein Text ins Kernkorpus integriert, der auch im Trainingskorpus enthalten ist (Hölderlins *Hyperion Fragment*). Dadurch ergibt sich einerseits ein realistisches Szenario der Anschlussnutzung, da bei nachgenutzten Tools nicht immer sichergestellt werden kann, dass Trainings- und Forschungsdaten sich nicht überschneiden. Auf der anderen Seite konnten so eventuell dadurch entstehende Verzerrungen im Blick behalten werden.

Der RIW bezieht sich auf den dieser Arbeit zu Grunde liegenden relativ weit gefassten Raumbegriff. Die Auswertung der literaturwissenschaftlichen Forschung zum Thema Raum hat gezeigt, dass Raum aber sehr unterschiedlich (eng) definiert werden kann. In einer umfassenden Studie, die es sich zum Ziel gesetzt hat, zu zeigen, welchen Stellenwert die Referenzierung von Rauminformationen in literarischen Texten hat, reicht es nicht, lediglich den eigenen, eher weit gefassten Raumbegriff, zu untersuchen. Darum wird für alle Texte des Kernkorpus auch noch der Anteil der Ortsnennungen im jeweiligen Roman berechnet

$$OIW = \frac{Anzahl\ Orte}{Gesamtwortzahl\ des\ Textes} \tag{7.2}$$

Auf diese Weise wird das engste mit dem hier entwickelten Modell operationalisierbare Raumverständnis ebenfalls in Form eines Indexwertes betrachtet. Dieser zweite Wert wird Orts-Index-Wert genannt (OIW) und die Höhe dieser Werte wird mit denen der RIWs abgeglichen. Auf ähnliche Weise und basierend auf den Grundannahmen, die der Formel des RIWs zu Grunde liegen, werden noch zwei weitere Indexwerte berechnet, der Relationen-Index-Wert (RelIW) und der Raumthemen-Index-Wert (ThIW). Beide Werte werden analog zu RIW und OIW berechnet. Im RelIW werden allerdings zwei Kategorien zusammengefasst, nämlich die Relationen und die relationalen Verben

$$RelIW = \frac{Anzahl\ Relationen + Anzahl\ relationale\ Verben \times 0,8}{Gesamtwortzahl\ des\ Textes} \tag{7.3}$$

Dem ThIW liegt die Kategorie der Raumthemen zu Grunde

$$ThIW = \frac{Anzahl\ zu\ Raumthemen\ gehoeriger\ Woerter \times 0,7}{Gesamtwortzahl\ des\ Textes} \tag{7.4}$$

Die vier Indexwerte werden pro Text berechnet und dann einzeln und zusammen diachron über das gesamte Kernkorpus verglichen. Wie in Kapitel 9 gezeigt werden wird, kann über die unterschiedlichen Teilaspekte eine Cross-Validierung von Thesen zur Bedeutung literarischen Raumes im diachronen Zeitverlauf vorgenommen werden. Berechnung, Auswertung und Analyse der Indexwerte schließen in Kapitel 9 die nun zuerst folgende Interpretation der Raum-Annotationen im Kernkorpus ab.

Teil III

Analysen

Check for updates

8 Orte und Räume in Romanen

Zur Auswertung der mit dem Raum-Classifier und dem StanfordNER automatisch erstellten Annotationen wurden die annotierten Texte der Teilkorpora in CATMA hochgeladen. Hier wurden Abfragen aller Raumkategorien in allen Dokumenten erstellt und ausgewertet.[1] Dabei wurden zunächst die jahrhundertspezifischen Teilkorpora betrachtet und anschließend das gesamte Kernkorpus. Am Ende wurden drei Perspektiven entwickelt, um die gefundenen Phänomene als Indikatoren der Raumdarstellung auf jeweils einer von drei Ebenen der Korpusbetrachtung zu klassifizieren:

1. Allgemeine Tendenzen literarischer Raumdarstellung und deren diachrone Entwicklung (Ebene des gesamten Kernkorpus);
2. Phänomene, die nur in einem Teilkorpus vorkamen (Ebene der Teilkorpora und damit der Literatur bestimmter Jahrhunderte);
3. Einzeltextphänomene (Ebene des einzelnen Romans).

Die Betrachtung allgemeiner Tendenzen literarischer Raumdarstellung erfolgt in insgesamt sechs beispielhaften thematischen Interpretationslinien (vgl. Abschnitt 8.1). Diese sind zum Teil auf einzelne Kategorien des Raummodells fokus-

[1] Alle hier ausgewerteten CATMA-Abfragen sowie ggf. deren Aufbereitung durch Klassifizierung oder Zusammenfassung habe ich in ein öffentliches GitHub-Repository hochgeladen, und über Zenodo zum Download bereit gestellt (Schumacher, 2021a). Verweise auf Unterordner stehen in den folgenden Kapiteln direkt bei der jeweiligen Analyse.

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-662-66035-5_8.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_8

siert, können aber auch kategorienübergreifend angelegt sein. Besondere Berücksichtigung fand bei den Auswertungen allgemeiner Tendenzen literarischer Raumdarstellung die Anschlussfähigkeit an die bisherige literaturwissenschaftliche Forschung. Die Phänomene, die nur in einzelnen Teilkorpora vorkamen, sind hauptsächlich raumthematisch. Hier werden in Abschnitt 8.2 jeweils pro Teilkorpus unterschiedlich viele Spezifika aufgezeigt, die in Form von explorativen, datenzentrierten Untersuchungen betrachtet werden. Das in Abschnitt 8.3 beschriebene Einzeltextphänomen, das bei der Datenauswertung dieser Studie entdeckt wurde, bezieht sich dagegen wieder auf unterschiedliche Kategorien des hier angewendeten Modells literarischen Raumes. Ausreißer-Phänomene wie dieses müssen unabhängig von der Vorgehensweise, sei diese nun auf Anschlussfähigkeit an bisherige literaturwissenschaftliche Forschung ausgelegt oder eher auf die Datenauswertung fokussiert, im Blick behalten werden, um keine Verzerrungen der Gesamtauswertung zu verursachen.

Zur Cross-Validierung der Analysen wurden diese außerdem zuerst mit einer Vorversion des Classifiers durchgeführt, die sich in einem Punkt von der Endversion unterschied. Die Metaphern wurden im Trainingskorpus gelöscht anstatt wie bei der finalen Version in die anderen Kategorien umsortiert zu werden. Diese Version des Classifiers erreicht eine um rund ein Prozent schlechtere Erkennung (74,77%). Durch den Vergleich dieser beiden Analyse-Durchläufe konnte festgestellt werden, welche Phänomene sich als besonders stichhaltig erweisen. Hat der Vergleich entscheidende Abweichungen aufgezeigt, so wird das bei der jeweiligen Teilanalyse im Folgenden explizit benannt.

Die beispielhaft angeführten Interpretationslinien dienen als Proof-of-Concepts und zeigen vor allem, inwiefern die hier entwickelte Methodik zur automatisierten Annotation von Indikatoren für Raumdarstellungen in literarischen Texten dazu dienen kann, Einstiegspunkte in die Betrachtung unterschiedlicher, literaturwissenschaftlich relevanter Phänomene zu finden. Dazu gehört auch, dass an dieser Stelle die entwickelte und angewandte Kernmethode der automatischen Annotation sinnvoll mit anderen Methoden der Digital Humanities verknüpft wird. Auf diese Weise werden Mixed-Methods-Ansätze gezeigt, die neben der CRF-Klassifikation auch Netzwerkanalyse, Aufbau, Abfrage und Visualisierung von Daten in einer relationalen Datenbank und Topic Modeling nutzen. Die insgesamt elf Ansatzpunkte zur genaueren Untersuchung literarischen Raumes können hier nicht vollumfänglich abgehandelt werden. In Anbetracht der Tatsache, dass nicht jeder gefundene Ansatz auf gleiche Weise anschlussfähig an traditionellere Methoden der Literaturwissenschaft ist, wird aber exemplarisch gezeigt, inwiefern die Auswertung der Daten des Raum-Classifiers An- und Rückschlüsse auf analoge, aktuelle literaturwissenschaftliche Forschung erlaubt. Die Methode der automatischen Klassifikation von

Indikatoren narrativer Raumdarstellungen wird also einerseits auf ihre Kombinierbarkeit mit anderen Methoden aus dem Bereich der Digital Humanities geprüft und andererseits auf die Möglichkeit, die gefundenen Phänomene mit Erkenntnissen aus den analog arbeitenden Literaturwissenschaften zusammenzuführen.

8.1 Perspektive: Gesamtkorpus

Im gesamten Korpus wurden vom Raum-Classifier 1.406.802 Annotationen eingefügt. Davon fallen 349.047 Annotationen auf das Teilkorpus des 18. Jahrhunderts, 425.265 auf das Teilkorpus des 19. Jahrhunderts, 307.815 auf das Teilkorpus des 20. Jahrhunderts und 313.513 auf das Teilkorpus des 21. Jahrhunderts (vgl. Tabelle 8.1). Abgesehen vom Teilkorpus des 19. Jahrhunderts ähneln sich die Annotationszahlen erstaunlich stark. Dass die Zahlen im 19. Jahrhundert im Vergleich mit den anderen Teilkorpora herausfallen, hängt damit zusammen, dass in diesem Jahrhundert sehr unterschiedlich mit Raumdarstellung umgegangen wurde und es einige Ausreißer-Texte gibt, bei denen der Anteil der Indikatoren für narrative Raumdarstellungen relativ hoch ist, wie die Auswertung der RIWs in Kapitel 9 zeigt. Durchschnittlich wurden in jedem der 100 Romane 14.068,02 Wörter als Teil von Raumdarstellungen annotiert. Das klingt zunächst einmal recht viel dafür, dass die Darstellung von Raum in literarischen Texten gegenüber anderen Phänomenen eher untergeordnet sein soll, aber natürlich schwanken die RIWs der einzelnen Texte stark, wie in der Gesamtübersicht noch gezeigt werden wird.

Tabelle 8.1 Annotationen pro Korpus und Kategorie

Kategorie	18. Jh	19. Jh	20. Jh	21. Jh	Gesamt
Relation	108.212	125.170	88.931	92.568	419.148
Relationales Verb	92.955	122.284	87.833	82.835	385.003
Raumhinweis	85.555	96.188	76.055	79.943	341.654
Ort	42.147	55.607	36.550	38.836	175.182
Raumthema	10.820	14.112	9.047	10.804	45.073
Raumbeschreibung	9.333	11.904	9.410	8.502	40.642
Gesamt	349.047	425.265	307.851	313.513	1.406.802
Tokens im Korpus	3.440.069	3.262.782	2.265.519	2.384.495	11.352.865

Die Auswertung der Annotationen des Raum-Classifiers hat gezeigt, dass die Rangfolge der Raumkategorien über das gesamte Kernkorpus hinweg fast kon-

stant bleibt (vgl. Tabelle 8.1). Immer sind Relationen auf dem ersten, relationale Verben auf dem zweiten Platz. Auf dem dritten Platz stehen in allen vier Teilkorpora Raumhinweise. Orte sind mit einem relativ großen Abstand immer auf Platz vier, Raumthemen meistens auf Platz fünf und Raumbeschreibungen fast immer auf Platz sechs. Nur im Teilkorpus des 20. Jahrhunderts gibt es eine Abweichung. Hier wurden etwas häufiger Raumbeschreibungen als Raumthemen annotiert. Zwischen Orten und Raumthemen (bzw. Raumbeschreibungen im 20. Jahrhundert) gibt es wieder einen recht großen Abstand, während Raumthemen und Raumbeschreibungen in ähnlicher Anzahl annotiert wurden. Die relationalen Raumausdrücke sind also quantitativ mit Abstand am bedeutsamsten für den literarischen Raum. Das Netz der Relationen ergibt sich, wie in Abschnitt 8.1.1 noch genauer ausgeführt werden wird, hauptsächlich aus Raumhinweisen als Knotenpunkten und Relationen und relationalen Verben als Indikatoren der Zwischenräume.

Raumhinweise sind aber nicht nur als Teil eines relationalen Raumnetzwerkes interessant, sondern dienen häufig auch der Verortung. Zwei zentrale Bereiche stehen einander hier gegenüber: Der Innenraum und der Außenraum. Beide werden häufig über Raumhinweise dargestellt oder impliziert (vgl. Abschnitt 8.1.2).

Für die Kategorie der Orte hat sich in erster Linie das Spannungsfeld zwischen dem Hier und dem Dort (vgl. Abschnitt 8.1.2) als quantitativ bedeutsam erwiesen. Das Setting von Erzähltexten wird viel seltener mit konkreten Ortsbenennungen – Städtenamen, Straßenbezeichnungen, Namen konkreter Etablissements usw. – dargestellt als mit diffusen und nur aus dem Kontext an einem bestimmten Punkt der Topografie eines Erzähltextes festzumachenden Bezeichnungen. Raumdarstellungen in literarischen Texten zeigen sich als Strukturphänomen, das eng mit dem der Figuren verknüpft ist. Diese können zwar auch unabhängig von Raumdarstellungen als absolute Entitäten betrachtet werden, Raumdarstellungen können aber vor allem im Kontext der häufig auftretenden Positionsbezeichnungen „hier“ und „dort“ nur im Zusammenhang mit den Figuren betrachtet werden, die eine Lokalisierung innerhalb der Geografie der storyworld erlauben. Dennoch können Figuren in literarischen Texten auch referenziert werden, ohne dass dabei Informationen zum Raum aufgerufen werden, z. B. wenn Eigenschaften einer Figur beschrieben werden.

Ein weiteres Phänomen, das sich auf das gesamte Kernkorpus erstreckt, ist die überragende quantitative Bedeutung des von Lotman ausgemachten literarischen Raumthemas der Öffnung und Schließung bzw. der Grenze (vgl. Lotman, 1973, 328–329 und Abschnitt 2.2.1 in diesem Buch). Nun ist dieses Thema bei Lotman auch sehr weit gefasst und jede Erwähnung einer Tür, eines Fensters, einer Wand oder einer Treppe (vgl. Lotman, 1973, 328–329), gehört für ihn zu diesem Raumthema, was das semantische Feld, das es bezeichnet, sehr groß macht. Dennoch gehört die

Grenzthematik eindeutig essentiell zur literarischen Raumdarstellung dazu, wie in Abschnitt 8.1.3 genauer beschrieben wird.

Eine deutliche Tendenz zunehmender Privatheit der in literarischen Texten dargestellten Räume, die sich im hier untersuchten Kernkorpus zeigt, bietet gute Anknüpfungspunkte dieser digital unterstützen Studie zu Betrachtungen aus der nicht digitalen Literaturwissenschaft. Die in den früheren Texten quantitativ bedeutsameren öffentlichen Räume fallen häufig in die Kategorie der Orte. Privatheit und damit verknüpfte Raumsymbole sind hingegen häufiger in Raumhinweisen impliziert. Auch Raumbeschreibungen wie die explizite Kennzeichnung eines Raumes als privat sind hier von Bedeutung. Wie die Entwicklung zunehmender Darstellung von Privatheit in literarischen Texten sich genau im Wortmaterial und in der Annotation des Raum-Classifiers zeigt, ist in Abschnitt 8.1.4 dokumentiert.

Von im diachronen Verlauf eher abnehmender Bedeutung sind die Begriffe „Welt" und „Himmel", die beide in die Kategorie der Orte fallen. Beide Begriffe vereint aber auch die Tatsache, dass sie oft metaphorisch verwendet werden oder Teil längerer metaphorischer Phrasen und Sprachbilder sind. In Abschnitt 8.1.5 wird die Manifestation ihrer abnehmenden Bedeutung im Kernkorpus und – mit Hilfe von Abfragen in der relationalen Metaphern-Datenbank *laRa* – ihre Bedeutung für die Raummetaphorik in literarischen Texten betrachtet.

8.1.1 Relationaler Raum in literarischen Texten

Die literarische Darstellung von Raum – das zeigt die Auswertung der automatischen Annotationen ganz deutlich – wird hauptsächlich über Relationen referenziert und entspricht damit einem Konzept, das u. a. von Leibniz vertreten und im 18. Jahrhundert dadurch auch in der Forschung stark diskutiert wurde (siehe Abschnitt 2.3.2). Insgesamt wurden im Kernkorpus 5.217 Types und 419.148 Tokens als Relation annotiert.[2] Das heißt, dass fast 2/3 aller mit dem Raum-Classifier erstellten Annotationen in diese Kategorie fallen. Die zehn am häufigsten annotierten Wörter sind „in", „auf", „an", „aus", „im", „zu", „vor", „von", „über" und „nach". „In" ist allerdings mit 81.198 Annotationen mit großem Abstand am häufigsten (am zweithäufigsten ist mit 43.819 „auf"). Die im gesamten Korpus am zweithäufigsten annotierte Kategorie ist die der relationalen Verben. Insgesamt wurden 385.003 Tokens annotiert. Es fallen also – nimmt man die Anzahl der Relationen und der relationa-

[2] Eine Tabelle mit den Zahlen der annotierten Types und Tokens in den Teilkorpora und im gesamten Kernkorpus befindet sich im GitHub-Repository zu dieser Arbeit im Ordner „Datenbasis der Analysen" (Schumacher, 2021a). Zu jeder Raumkategorie gibt es einen Unterordner, in dem die Abfragen der vier Teilkorpora und des gesamten Kernkorpus zu finden sind.

len Verben zusammen – mehr als die Hälfte aller Annotationen in die relationalen Kategorien des hier entwickelten Raummodells. Mit 17.871 Types sind relationale Verben aber viel diverser als Relationen und eignen sich auch besser, um, darauf aufbauend, Interpretationshypothesen aufzustellen.

Die am dritthäufigsten annotierte Kategorie ist mit 341.654 Annotationen die der Raumhinweise. Auch diese Kategorie trägt häufig dazu bei, einen relationalen Raum zu entfalten, denn relationale Verben und Relationen bezeichnen meist Zwischenräume oder die Überbrückung derselben. Diese Zwischenräume entfalten sich zwischen Figuren und / oder zwischen Figuren und Objekten. Referenzen darauf müssen nicht immer Hinweise auf den Raum sein, Raumhinweise bezeichnen aber häufig Objekte, Figuren oder einzelne Teile und Teilbereiche von Objekten oder Figuren. Die Wortvielfalt der Raumhinweise ist hoch, die Annotationen verteilen sich auf 40.295 Wörter. Nicht immer stehen Raumhinweise in engem Zusammenhang mit Relationen, aber sehr häufig. Dass Raumhinweise und Relationen und relationale Verben tatsächlich im Satzzusammenhang oft direkt nebeneinander stehen, zeigt eine Netzwerkanalyse der Raumkategorien in den vier Teilkorpora (vgl. Abbildung 8.1).

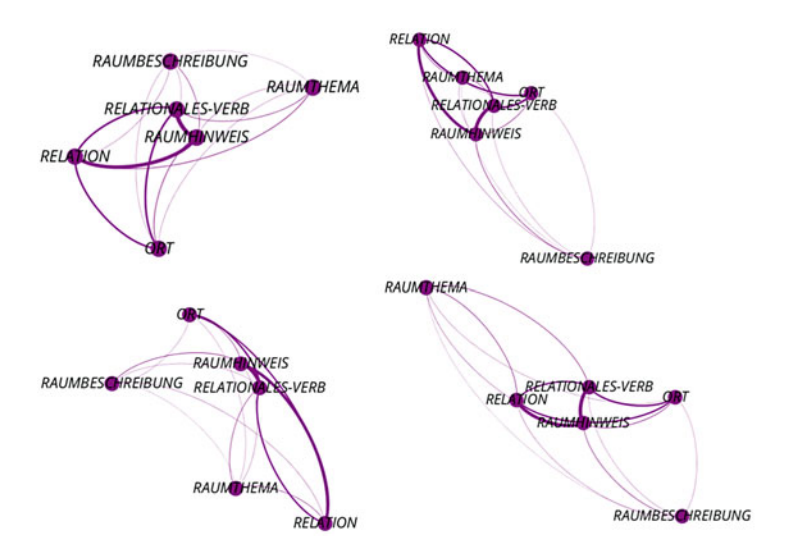


Abbildung 8.1 Netzwerkvisualisierung der Raumkategorien

Die Netzwerkvisualisierungen der Annotationen in den vier Teilkorpora geben einen Eindruck davon, wie häufig Wörter, die in den unterschiedlichen Raumkategorien annotiert wurden, direkt nebeneinander stehen (Visualisierung der *first neighbours*). Als Visualisierungstool wurde hier *Gephi* eingesetzt, eine Software zur Netzwerkerstellung, die unterschiedliche Layouts für Netzwerke anbietet, die auf unterschiedlichen physikalischen Simulationen beruhen (Schumacher, 2020b, §68ff). Die automatisch annotierten Daten, die zur Analyse in CATMA hochgeladen worden waren, wurden mithilfe des Python-Packages GitMA (Vauth et al., 2022) für die Netzwerkanalyse aufbereitet. Grundidee ist, dass zwischen den Knoten Anziehungskräfte bestehen, die Kanten diese aber wie Sprungfedern auseinander drücken. Je nachdem, wie viele Verbindungen ein Knoten zu anderen hat und wie stark diese sind, ergibt sich für eine bestimmte Netzwerkstruktur ein Gleichgewicht. Das hier ausgewählte Layout ist OpenOrd, ein Layout, dass die schwächeren Verbindungen in der Simulation abschneidet und sie so in der Visualisierung von den anderen Knoten und Kanten abrückt (Schumacher, 2020b, §79). Das führt dazu, dass die starken Verbindungen sehr gut sichtbar werden, täuscht aber in diesem Fall etwas, was die Konstellationen im Netzwerk angeht. Da alle Knoten Verbindungen zu allen anderen aufweisen, sind die Kanten eigentlich alle gleich lang und nur unterschiedlich dick. Dadurch, dass die dickeren Kanten hier auch kürzer dargestellt sind, zeigt sich aber sehr schön, dass die Wörter der Kategorien der relationalen Verben und der Raumhinweise in allen vier Teilkorpora am häufigsten direkt nebeneinander stehen, wie z. B. in der Phrase „jemandem die Hand [Raumhinweis] reichen [relationales Verb]“. Relationen stehen dann am nächst häufigsten neben Raumhinweisen (z. B. „zu [Relation] ihr [Raumhinweis]“). Relationen und relationale Verben (z. B. „verborgen [relationales Verb] in [Relation]“) stehen in etwa gleich häufig nebeneinander wie Orte und relationale Verben (z. B. „die Wohnung [Ort] betreten [relationales Verb]“) und die Wörter der Kategorien Raumthema und Raumbeschreibung (z. B. „dicke [Raumbeschreibung] Mauern [Raumthema]“) stehen weit seltener direkt neben denen der anderen Kategorien.[3] Es gibt aber keine Kategorie, deren Wörter niemals direkt neben denen der anderen stehen, die Vernetzungsgrade sind in allen vier Netzwerken maximal (nämlich bei 1 – jeder Knotenpunkt weist Relationen zu allen anderen auf). Die vergleichsweise häufige direkte Nachbarschaft der Wörter der Kategorien relationale Verben, Raumhinweise und Relationen deutet auf die grundsätzlich relationale Beschaffenheit literarischer Raumdarstellungen hin:

[3] Raumthemen stehen hier zwar ebenfalls etwas für sich in den Netzwerken, diese Kategorie bildet aber einen Sonderfall. Da unter Raumthemen auch Wörter fallen, die zu anderen Kategorien zählen würden, wenn sie nicht zu einem Raumthema beitragen würden (wie z. B. das relationale Verb „reisen“), muss die Ferne zu den anderen Raumkategorien hier etwas relativiert betrachtet werden.

Subjekte und Objekte, die häufig unter die Kategorie Raumhinweise fallen, überbrücken, wenn sie mit relationalen Verben zusammen stehen, Distanzen. Relationen drücken genau diese Distanzen aus. Der literarische Raum ergibt sich aus diesem Zusammenspiel, in dem Distanzen aufgebaut und überbrückt werden (oder auch bestehen bleiben). Orte bezeichnen zwar das Setting, tragen aber weniger zu diesem Zusammenspiel bei, ebenso wie Raumbeschreibungen. Der relationale Raum, der in literarischen Texten entworfen wird, ist also überwiegend von Subjekt-Subjekt und Subjekt-Objekt-Relationen geprägt und weit seltener von Distanzen, die sich zwischen Orten erstrecken. Es ist also nicht hauptsächlich ein topografisch-relationaler Raum, sondern eher ein subjekt- bzw. objektorientierter.

Wenn man nun alle Annotationen der so oft nebeneinander stehenden Kategorien zusammenrechnet, die maßgeblich zum relationalen Raum beitragen, so stellt man fest, dass sie rund 80% (genau 81,44%) aller Annotationen im Kernkorpus ausmachen. Auch wenn nicht jeder Raumhinweis zu einem relationalen Raumgefüge beitragen muss, so ist dieser Wert doch so hoch, dass der Schluss nahe liegt, dass der Raum, der in literarischen Texten dargestellt wird, hauptsächlich relational ist.

Zur Interpretation der Kategorie „Relation“
Wörter, die Relationen ausdrücken, also z. B. Deiktika, Bezeichnungen von Zwischenräumen oder von Nähe und Ferne werden am häufigsten dazu genutzt, Informationen zum narrativen Raum zu referenzieren. Abgesehen von dieser These bietet diese Kategorie aber kaum Interpretationspotential. Dass „in“ hier mit Abstand am häufigsten (81.198 Mal) annotiert wurde, sagt, in Anbetracht der Tatsache, dass die Kontexte höchst variabel sind, nicht viel aus. Wie wenig sich die relative Verwendung der hier annotierten Wörter unterscheidet, zeigt sich im Vergleich der Teilkorpora miteinander. Die Rangfolge der ersten vier Wörter dieser Kategorie („in“, „auf“, „an“, „aus“) ist in allen vier Teilkorpora genau gleich.[4] Erst danach ergeben sich jahrhundertspezifische Besonderheiten, die sich allerdings erst einmal nur in minimalen Abweichungen der Rangfolge zeigen. Im Teilkorpus des 18. Jahrhunderts ist z. B. „zu“ auf Rang fünf und „im“ auf Rang acht. Letzteres ist aber im Teilkorpus des 20. Jahrhunderts auf Rang fünf. Was sagt das im Hinblick auf eine literaturwissenschaftliche Analyse aus? In erster Linie wohl, dass bei Distant-Reading-Analysen wie der vorliegenden Vorsicht vor zu schnellen Schlüssen geboten ist. Betrachtet man nämlich genauer, in welchen Zusammenhängen z. B. das Wort „in“ im Teilkorpus des 18. Jahrhunderts stehen kann, dann sieht man vor

[4] Eine Übersicht über die genauen Zahlenwerte der einzelnen als „Relation“ annotierten Wörter, für jedes Teilkorpus und das gesamte Kernkorpus, befindet sich im GitHub-Repository zu dieser Arbeit im Ordner „Datenbasis der Analysen / Relationen“ (Schumacher, 2021a).

allem, dass dieses Wort in sehr, sehr unterschiedlichen Kontexten steht.[5] Die Häufung deutet also keineswegs darauf hin, dass sich Vieles in einem klar definierten Raum abspielt, wie man intuitiv vermuten könnte. Stattdessen zeigen häufig vorkommende Wendungen wie „in die Welt hinaus“, „in allen Gassen“, „in das düstere Tal“, dass „in“ auch in Verbindung mit sehr unkonkreten Räumen stehen kann. Von Ländern über Städte bis zu Zimmern und letztendlich auch bis in Möbelstücke oder, noch kleiner, Kisten, sind Verortungen im Inneren und Äußeren mit dem Wort „in“ gekennzeichnet. Darüber hinaus können auch gerichtete Bewegungen, Blicke oder Rufe, die mit „in“ gekennzeichnet werden (z. B. im Satz „und hier sah er in das Dörfchen“ in Teilkorpus 18: Jean Paul *Hesperus oder 45 Hundsposttage*). Tatsächlich kann also keine inhaltlich ausgerichtete, belastbare These darüber aufgestellt werden, welche Art von Relationen in einer Zeitspanne mit welchen raumspezifischen Motiven einhergeht und besonders häufig ist. Hinzu kommt, dass das Verhältnis der Häufigkeiten der Wörter, die als Relation annotiert wurden, in allen vier Teilkorpora so ähnlich ist, dass die Vermutung nahe liegt, dass hier weniger eine Korpus- als vielmehr eine Sprachspezifik vorliegt. „In“ scheint einfach insgesamt die am häufigsten zur Raumdarstellung genutzte Relation zu sein.

Eine Frage, die sich bei Betrachtung der Annotationen der Kategorie „Relation“ geradezu aufdrängt, ist, ob die Wörter, die in diese Kategorie fallen, nicht über einen Listenabgleich hätten annotiert werden können. Tatsächlich ist das Wortmaterial der in dieser Kategorie annotierten Wörter sehr klar umrissen. Allerdings sind selbst diese kleinen und unscheinbaren Wörter nicht frei von Ambiguitäten und können in unterschiedlichen Kontexten sehr unterschiedlich genutzt werden. In dem Satz „wenn er in dieser Stimmung war, war einfach nichts mit ihm anzufangen“ bezeichnet das Wort „in“ gar keine Relation. Ein kontextsensitives Tool wie der auf Machine Learning basierende Raum-Classifier kann diese Unterschiede in der Verwendung der Wörter im Satzzusammenhang berücksichtigen. Dennoch kann der hier gewählte Ansatz auch mit Formen des Listenabgleichs zusammengebracht werden, wie in den folgenden Kapiteln noch gezeigt werden wird.

Zur Interpretation der Kategorie „relationale Verben“

Die Kategorie der relationalen Verben bietet deutlich mehr Interpretationspotential als die der Relationen. Dass im gesamten Kernkorpus mit „sehen“ ein Ausdruck optischer Wahrnehmung mit 8.790 Annotationen allein von der Form „sah“ am häufigsten ist, ist sehr bezeichnend. Nicht nur, dass die von Husserl (vgl. Abschnitt 2.3.2)

[5] Eine Keyword-in-Kontext-Tabelle dieses Wortes befindet sich im GitHub-Repository der Forschungsdaten dieser Arbeit unter „Datenbasis der Analysen / Relationen / Annotationen des Wortes in Kontexte 18Jh.csv“ (Schumacher, 2021a).

beschriebenen Sichtachsen auch für literarische Texte bedeutsam sind, drückt sich hier aus. Es zeigt sich auch, dass Lesende zwar – anders als beim Werk der bildenden Kunst, um auf Lessing (vgl. Kapitel 1) zurückzukommen – nicht direkt auf eine räumlich manifestierte Szenerie schauen, aber dass der Blick darauf durchaus dargestellt wird, und zwar vermittelt durch Erzählinstanz und / oder Figuren. Durch die Augen derer, die die storyworld beleben, wird auch der Blick der Lesenden gelenkt. Quantitativ weniger dominant sind Verben der Bewegung wie „kommen", „gehen", „stehen" oder „nehmen"[6], also Verben, die eine zeitliche Abfolge implizieren. Obwohl das Lesen an sich eine Tätigkeit ist, die in zeitlicher Linearität abläuft, muss das Dargestellte also nicht immer analog dazu erfasst werden. In der Beschreibung dessen, was eine Figur sieht, kann durchaus räumliche Plastizität erreicht werden.

In allen vier Teilkorpora (18.–21. Jahrhundert) ist „sah" das am häufigsten annotierte Wort der Kategorie „relationale Verben". Im Teilkorpus des 18. Jahrhunderts wurde „sah" 2.082 Mal annotiert, im Teilkorpus des 19. Jahrhunderts 2.705 Mal, im Teilkorpus des 20. Jahrhunderts 2.389 Mal und im Teilkorpus des 21. Jahrhunderts 1.614 Mal. Gruppiert man andere Formen des Verbs „sehen" hinzu, so zeigt sich die quantitative Bedeutung noch deutlicher. Im vierten, dem Teilkorpus des 21. Jahrhunderts, ist mit 1.547 Annotationen das Wort „an" als Teil eines relationalen Verbs fast ebenso häufig wie „sah". Im gesamten Kernkorpus sind also Sicht- und Blickachsen von überragender Häufigkeit in der Darstellung des Raumes. Im Teilkorpus des 21. Jahrhunderts kommt durch das häufig im Zusammenhang mit optischer Wahrnehmung auftauchende „an" (z. B. in „ansehen", „anblicken" oder „anschauen") zu einer stärkeren Richtung dieser Achsen.

Eine genauere Analyse der Zusammenhänge des am häufigsten mit dieser Kategorie annotierten Wortes „sah" hat gezeigt, dass die Kontexte in vier Gruppen zusammengefasst werden können:

1. Blick auf Figuren
2. Blickachsen zwischen Figuren
3. Blickachsen auf Szenerie und Setting
4. Blick auf Lebensansichten der Figuren

Sehr häufig schauen Figuren sich aber auch ein- oder wechselseitig an, sodass sich zwischen ihnen eine Blick-Relation ergibt. Beispiel für das einseitige ansehen, ist

[6] Die genauen Auflistungen der Wörter, die als „relationales Verb" annotiert wurden pro Teilkorpus und im gesamten Kernkorpus sind im GitHub-Repository zu dieser Arbeit im Ordner „Datenbasis der Analysen / relationales Verb" zu finden (Schumacher, 2021a).

die Passage „. . . sie sah ihren Vater vor sich stehen“ (Teilkorpus 18: Huber *Die Familie Seldorf*) für beidseitiges anschauen ist „Die Mädchen sahen einander verlegen an“ (Teilkorpus 18: Vulpius *Rinaldo Rinaldini*) ein Beispiel. Schilderungen des Aussehens von Figuren lassen diese beim Lesen plastisch wirken und verleihen den beschriebenen Charakteren eine optische Komponente. „Ich sah ihn in dem Glanze seiner wirklich herrlichen Schönheit“ (Teilkorpus 19: *Spielhagen Problematische Naturen*) oder „Der Gast sah übel aus“ (Teilkorpus 19: Franzos *Leib Pfefferkuchen*) sind nur zwei von zahlreichen Beispielen für diese Verwendung. Besonders interessant ist auch, dass über die Blickachse auf die Optik einer Figur auch ein Eindringen in deren Innenwelt geschildert werden kann: „man sah wie tief ihn der Gegenstand und das Scheitern vieler Erwartungen und gehegten Ansichten berührte.“ (Teilkorpus 19: Goedsche *Sebastopol*), oder „man sah es ihm ordentlich an, welch’ Vergnügen ihm ein Kampf mache“ (Teilkorpus 19: Gerstäcker *Flusspiraten des Mississippi*) sind zwei relativ willkürlich aus der Menge ausgewählte Beispiele für ein solches Eindringen der Blickachse in eine Figur. Dieses Eindringen kann natürlich auch bewusst abgeschnitten werden, wie die Textzeile „Seine Wangen waren erglüht. Er sah zu Boden.“ (Teilkorpus 19: Stendhal *Rot und Schwarz*) belegt.

> Es war der Hohenstein. Sie kletterten über eine gleiche Holztreppe hinauf. Von ihm sah Witiko die Berge des böhmischen Landes höher und breiter als von den Sesseln, auch sah er neue Waldlehnen emporstehen, aber weiter nichts von dem inneren Lande. (Teilkorpus 19: Stifter *Witiko*)

In Passagen wie dieser von Stifter oder auch in kürzeren Phrasen wie „[v]or ihnen sah man über die Wipfel hinweg zur Ebene“ (Teilkorpus 19: Spielhagen *Problematische Naturen*) werden Blicke auf Szenerie und Setting dargestellt. Diese Art von Blickachsen führen häufig über Grenzen bzw. Begrenzungen hinweg. Figuren schauen von innen aus dem Fenster oder aus dem Haus oder von außen hinein.

Seltener sind Darstellungen von Lebensansichten wie z. B. „[s]ie sah alles für ein Verhängnis an“ (Teilkorpus 18: Gellert *Das Leben der schwedischen Gräfin von G.*). Diese metaphorische Verwendung des Wortes „ansehen“ fällt nicht in die Kategorie der „relationalen Verben“, sondern in die der „Raummetaphern“. Sie ist darum hier nicht von zentralem Interesse. Da der Raum-Classifier in der finalen Version Raummetaphern nicht berücksichtigt, muss dieser Kontext aber durchaus mit einbezogen werden, wenn mit Hilfe dieses Tools eine Analyse des Verbs „sehen“ und verwandter Formen durchgeführt wird.

Die Kontexte des Sehens sind insgesamt vor allem blickführend. Szenerien werden eröffnet, wobei zum Teil Grenzen durchbrochen werden (von innen nach außen

Tabelle 8.2 Am häufigsten als „relationales Verb“ annotierte Wörter in den vier Teilkorpora

Rang	18. Jh	19. Jh	20. Jh	21. Jh
1	sah	sah	sah	sah
2	kam	rief	kam	an
3	sehen	ging	ging	sehen
4	an	sehen	stand	ging
5	kommen	stand	an	gehen
6	ging	kam	sehen	kam
7	rief	kommen	kommen	stand
8	gehen	an	rief	aus
9	nahm	gehen	hörte	auf
10	stand	gekommen	gehen	sieht

oder andersherum), Figuren werden betrachtet und Blickachsen zwischen Figuren als verbindende Elemente gestaltet. Auch hierbei können Grenzen gewahrt, gesetzt oder überschritten werden, nämlich die Grenzen zwischen Außen- und Innenwelt einer Figur. Gemeinsam mit der Figur (oder auch der Erzählinstanz), deren Blicke gerade verfolgt werden, sieht die lesende Person auf diese Weise den Raum der storyworld mit.

In zwei der vier Teilkorpora stehen an zweiter Stelle der am häufigsten als relationale Verben annotierten Wörter Bewegungsverben (vgl. Tabelle 8.2). „Kam“ ist im Teilkorpus des 18. und 20. Jahrhunderts am zweithäufigsten. Im Teilkorpus des 21. Jahrhunderts steht nach „an“, „sah“ und „sehen“ an vierthäufigster Stelle das Bewegungsverb „ging“. „Kam“ und „stand“ folgen auf Platz sechs und sieben der am häufigsten annotierten relationalen Verben (auf Platz fünf steht „gehen“). Auch in den anderen Teilkorpora sind Formen des Wortes „gehen“ unter den zehn am häufigsten in dieser Kategorie annotierten Wörter. Es ergibt sich ein insgesamt relativ homogenes Bild, in dem Wahrnehmungsverben, vor allem „sehen“ und gerichtete Bewegungsverben, die zu etwas hin- oder von etwas wegführen, dominieren. Im 21. Jahrhundert nehmen die Richtungen der Blickachsen, die schon bei Betrachtung der anderen Teilkorpora als bedeutsam für die Raumdarstellung identifiziert wurden, an Wichtigkeit zu. Neben Formen des Verbs „sehen“ sind hier vor allem Bewegungsverben häufig, wie z. B. „gehen“, „stehen“ und „kommen“ (siehe Tabelle 8.2). Andere Ausdrücke von Sinneswahrnehmungen wie „hören“ und auch das häufig eng damit verknüpfte „rufen“ sind seltener oder nehmen im diachronen Verlauf an Häufigkeit ab (wie z. B. der Fall bei „rief“ – siehe Tabelle 8.2).

Sowohl die quantitative Bedeutung der Wahrnehmungsverben als auch die der Bewegungsverben für den relationalen literarischen Raum zeigen, dass hier Konzepte der in Kapitel 2 herausgearbeiteten Aspekte des Raumes implementiert wurden und sich darum auch in den Daten widerspiegeln. Die Blick- und Rufachsen nach Husserl wurden bereits erwähnt. Hinzu kommt aber auch das Konzept des Bewegungsraums als Handlungsraum nach de Certeau (vgl. Abschnitt 2.3.2) und das des sozialen Raumes nach Foucault (vgl. Abschnitt 2.3.3). Der zunächst zweidimensional erscheinende, netzwerkartige relationale Raum wird dadurch mehrdimensional. Neben Distanzen und Verknüpfungen werden hier auch Aspekte der zeitlichen Dimension wichtig. „Gehen" impliziert eine langsamere Bewegung als „rennen", während im relationalen Verb „kommen" Geschwindigkeit keine oder nur eine untergeordnete Rolle spielt. Die beispielhaft angeführten Verben „gehen" und „rennen" beinhalten aber auch eine Fortbewegungsform, die ohne technische Hilfsmittel auskommt und darum hier auch nicht als dem Raumthema der Reise zugehörig interpretiert wird. Die Distanzen, die mit dieser Bewegungsart überbrückt werden, sind in der Regel eher kurz als lang. Genau wie de Certeau (2006, 345) beschreibt, ist Raum hier „eine mehrdeutige Einheit". Das Konzept erweist sich ebenfalls als passend für die Sicht- und Rufachsen, da hier neben Distanzen, Zeitlichkeit und Blickführung z. B. auch Emotionen eine wichtige Rolle spielen. Bricht sich eine Emotion z. B. durch einen Ausruf Bahn und dringt somit von der Innen- in die Außenwelt, so kann hier – neben der Emotion selbst – auch noch positive und negative Aufladung derselben, sowie Heftigkeit eine Rolle spielen. Blickachsen, die unterbrochen werden, deuten ebenfalls auf eine emotionale Dimension hin, die hier zu einer direkten Konsequenz für den relativen Raum führt, indem eine Relation unterbrochen wird. Dass Raum „mit zahlreichen Qualitäten behaftet ist und möglicherweise auch voller Phantome steckt" thematisiert auch Foucault (2006, 319). Darüber hinaus spricht er dem Raum viele Qualitäten zu und beschreibt unter anderem Aspekte des individuellen und sozialen Raums des Innen- und Außenraums von Personen (vgl. Foucault, 2006, 319 und Abschnitt 2.3.3). Dieser Foucault'sche Handlungsraum zeigt sich immer wieder in den hier betrachteten Romanen. Es ist gerade diese Mehrdeutigkeit und Mehrdimensionalität des relationalen Raumes, die es nicht erlaubt, für literarische Raumdarstellungen ein statisches Konzept anzunehmen. In Abschnitt 2.3.4 wurde ein Exkurs in Richtung Archäologie gewagt, um zu zeigen, dass Raum auch mit Hilfe haptischer Elemente wie Erde und Stein, aus denen in der wissenschaftlichen Arbeit einzelne Objekte geborgen werden, betrachtet und zum Teil besser verstanden werden kann. Literarische Raumdarstellungen aber zeigen etwas anderes. Hier gilt es nicht wie in der Archäologie die Schichten von Wortmaterial zwischen den Raumausdrücken abzutragen, sodass eine feste, dem Text unterliegende Struktur zu Tage tritt. Zwar implementiert die Methode der automatischen Raum-Klassifikation

einen Teil dieser Idee und lässt die für die Raumdarstellung am häufigsten genutzten Ausdrücke deutlich hervortreten; in der Interpretation gilt es aber das, was gefunden wurde, durchaus im Kontext seiner Variabilität und Mehrdeutigkeit zu betrachten. Vor allem bei den dominierenden relationalen Aspekten von Raum wird hier ganz klar, dass dieser dynamisch konstruiert wird, und zwar nicht nur auf lineare Weise. Der relationale literarische Raum wird nicht im Handlungsverlauf aufgebaut und ist dann am Ende auf irgendeine Weise ein festes Konstrukt. Stattdessen wandelt sich der relationale literarische Raum in jedem Moment des Lesens, indem sich Relationen auf- und wieder abbauen, durchkreuzt, zerstört oder erzwungen werden. Der relationale literarische Raum ist also mehrdimensional (im Sinne von de Certeau (2006) und Foucault (2006)) und dynamisch, wie schon Klotz (1969) und Hoffmann (1978) ihn beschrieben haben. Hinzu kommt, dass die literarische Raumerfahrung auch Rückbezüge zu einer konkret-empirischen und zu einer chronotopischen (nach Bachtin, 2008) Ebene zulässt und somit in ein komplexes Gefüge eingebettet ist. Der Raum, der in einem literarischen Text referenziert wird, kann auf eine konkrete Geografie bezogen werden, ist aber immer auch in literarhistorische Zusammenhänge eingebunden, die vor allem auf soziale und kulturelle Aspekte des Raumes einwirken.

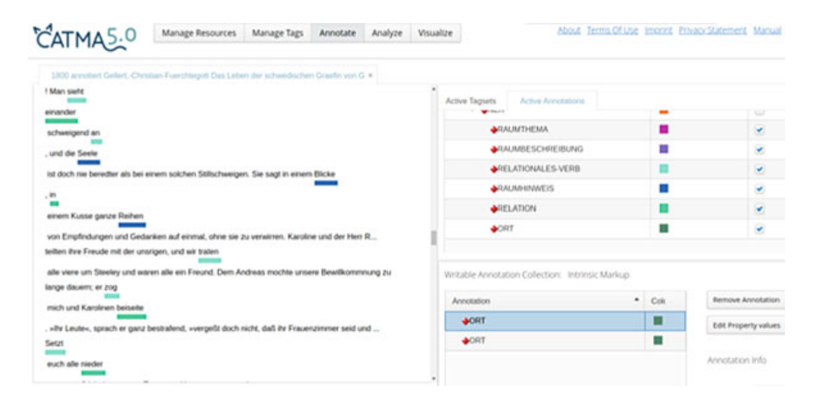


Abbildung 8.2 Passage aus Gellert *Das Leben der schwedischen Gräfin*

In der in Abbildung 8.2 illustrierten Passage zeigt sich deutlich diese Mehrdimensionalität und Dynamik und auch, welche Indikatoren davon vom Raum-Classifier identifiziert wurden. Durch das gegenseitige Ansehen wird eine Blickachse aufgebaut, die in erster Linie die Innenwelten der Figuren miteinander verbindet. Durch diese Relation werden Emotionen sichtbar, aber auch weitere ausgelöst. Es wird ein

Kreis um eine der Figuren gebildet, sodass eine Art kollektive Umarmung entsteht, mehr noch: Die fünf erwähnten Figuren werden zu einer einzigen Größe im Raum („waren alle ein Freund"). Dann greift eine andere Figur hinein und löst diese Konstellation wieder auf. Die Figur des Stanley erklärt in der in Abbildung 8.2 gezeigten Passage sein Verhalten durch gesellschaftliche Konventionen und bittet alle, sich zu setzen. Damit ist auch die Situation beendet und das räumliche Gefüge, das eben noch Bestand hatte, wird aufgelöst. Hier zeigt sich auch sehr schön, dass die Annotationen an einzelnen Größen als Indikatoren der Konstellation festgemacht sind. Die soziale und gesellschaftlich kulturelle (und damit auch zeitspezifische und also chronotopische) Ebene des dargestellten Raumes entfaltet sich aber erst, wenn die gesamte Textpassage betrachtet wird.[7]

Bei der Betrachtung der Bewegungsverben darf nicht vergessen werden, dass diese zum Teil auch mit immateriellen Dingen im Zusammenhang stehen. Das Verb „kommen" steht z. B. häufig in solchen Kontexten (wie im Teilkorpus 18 in Wezels *Herrmann und Ulrike* in den Passagen „die Unschuld kam an den Ort der Verführung" und „wenn sie vor mir vorüber gehen, kommen mir alle wie Zwerge vor" oder in Schnabels *Wunderliche Fata einiger Seefahrer*: „. . . da es nach der Hochzeit zur Untersuchung kam"). In der Regel wird aber wirklich Bewegung ausgedrückt. Sehr häufig ist „zurück-", „wieder-" und „ankommen", genauso wie „zu jemandem oder etwas (hinein-)kommen". Sehr ähnlich sind die Kontexte des Verbs „gehen". Auch hier gibt es einige wenige Wendungen, die nicht auf Bewegung hindeuten (z. B. „in Erfüllung gehen"). Meistens wird aber tatsächlich ausgedrückt, dass eine Figur sich auf einem Weg befindet, befunden hat oder befinden wird.

Etwas überraschend ist es, dass im Teilkorpus des 19. Jahrhunderts kein Bewegungsverb nach „sehen" am häufigsten als relationales Verb annotiert wurde, sondern mit 2.487 Annotationen „rief" (vgl. hierzu auch Tabelle 8.2). Betrachtet man

[7] Nantke kommt ebenfalls zu dem Schluss, dass verschiedene Raum-Formen nur dynamisch adäquat erfasst werden können (Nantke, 2017, 206), obwohl sich ihr Forschungsgegenstand in entscheidender Weise von dem für diese Studie gewählten Korpus unterschiedet: Mit der Analyse des Werkes von Kurt Schwitters ist es ihr möglich, eine komparatistische Betrachtung durchzuführen, die Aspekte der bildenden Kunst mit einbezieht. Obwohl sich das zu Grunde liegende Raumkonzept gegenstandsbedingt von dem in dieser Arbeit entwickelten Raummodell stark unterscheidet, kann die Dynamik als gegenstandsunabhängig bedeutsam ausgemacht werden. Dass diese auch für Texte wichtig ist, die Elemente der bildenden Kunst beinhalten, führt zurück zu Lessings Laokoon-These und legt nahe, dass zeitliche Aspekte für die bildende Kunst möglicherweise eine ebenso große Rolle spielen wie die in dieser Studie untersuchte Raumdarstellung für literarische Texte. Weitere komparatistische Studien, die sowohl bildende Kunst als auch Literatur einbeziehen und die digitale Methoden nutzen, um größere Korpora einzubeziehen, wären ein vielversprechender Anschluss.

die Kontexte des Verbs „rufen“ in diesem Teilkorpus genauer,[8] so stellt man fest, dass es auf zweierlei Weisen raumrelational fungiert. Einerseits rufen Figuren einander zu, um nicht allzu große Distanzen zu überbrücken. „Als sie das junge Mädchen erblickte, nickte sie freundlich und rief hinauf“ (Teilkorpus 19: Marlitt *Goldelse*) oder „Sie schlug eine der Gardinen zurück und rief hinaus ‚Bärbchen, hast du etwas auszurichten? Wir gehen spazieren und kommen zu deinem Schuh- und Hochzeitmacher!‘“ (Teilkorpus 19: Keller *Das Sinngedicht*) sind zwei Beispiele dieser Verwendung. Auf der anderen Seite bricht sich durch Rufen häufig eine innere Empfindung Bahn und dringt nach außen. „‚Das Horn! Das Horn!‘ rief er wild. ‚Unser Leben hängt von dem Pulver ab!‘“ (Teilkorpus 19: Goedsche *Sebastopol*), „‚Heilige! Verklärte!‘ rief jetzt dieser, außer sich vor unaussprechlicher Angst“ (Teilkorpus 19: Schopenhauer *Gabriele*) oder „‚Sprich du, Isidor, wenn ihr etwas zu sagen habt!‘ rief sie in leidenschaftlicher Vergessenheit dennoch lauter, als sie wollte.“ (Teilkorpus 19: Keller *Martin Salander*) sind drei Belege für eine solche Kontextualisierung des Verbs „rufen“. Die räumliche Komponente ist hier gewiss nicht vordergründig, denn eigentlich stehen die Emotionen im Fokus, und doch gibt es sie, da das Empfundene von innen nach außen dringt. Wahrnehmung und Ausdruck der Figuren sind also auch Grenzphänomene, bei denen etwas aus dem Körper herausdringen oder in anderen entdeckt werden kann. Ruf- und Blickachsen sind nicht nur potentielle Verbindungen zwischen äußerlich begrenzten Körpern, sondern auch zwischen Innen- und Außenwelten.

Im Teilkorpus des 18. Jahrhunderts wurde „rief“ zwar sehr viel seltener annotiert (1.094 Mal), steht aber in der Rangliste der häufigsten Wörter der Kategorie relationale Verben immerhin auf Platz sieben. Im Teilkorpus des 20. Jahrhunderts steht es auf Platz acht, in dem des 21. Jahrhunderts erst auf Platz 16. Während Sichtweiten also jahrhundertübergreifend von besonderer Bedeutung für literarische Raumdarstellung zu sein scheinen, sind Ruf-Distanzen und auch der Ausdruck von Empfindungen durch Ausrufe in den früheren Jahrhunderten – insbesondere im 19. – häufiger als in zeitgenössischer Literatur.

Auch die weiteren in dieser Kategorie annotierten Wörter tragen meist zu den drei oben genannten Aspekten bei; Szenerien werden eröffnet, Distanzen zwischen Figuren, und Figuren und Objekten werden durch Sichtachsen (seltener auch durch andere wahrnehmungsbezogene Verben wie „rufen“ / „hören“ oder „tasten“) oder Bewegungsverben überbrückt und Innen- und Außenwelten werden miteinander in Beziehung gesetzt. Dabei spielen Grenzphänomene zwischen diesen Welten eine

[8] Eine Keyword-in-Kontext-Tabelle dieses Wortes befindet sich im GitHub-Repository der Forschungsdaten dieser Arbeit unter „Datenbasis der Analysen / Relationales Verb / Annotationen des Wortes rief Kontexte 19Jh.csv“ (Schumacher, 2021a).

besondere Rolle. Der hier betrachtete relationale literarische Raum ist dynamisch und erlangt im jeweiligen Kontext Mehrdeutigkeit, bzw. erstreckt sich über mehrere Dimensionen, wie z. B. neben der räumlichen auch der kulturellen und gesellschaftlichen.

Zur Interpretation der Kategorie „Raumhinweise"
Die Auswertung der Kategorie „Raumhinweise" unterstützt die oben ausgeführten Beobachtungen zu den am häufigsten annotierten relationalen Verben. „Augen" wurden in dieser Kategorie am häufigsten annotiert (9.193 Annotationen) und auch „Blick" wurde oft mit dem „Raumhinweis"-Tag versehen (3.250 Mal). Sehr häufig sind Augen Teil der Beschreibung von Blickachsen, die Zwischenräume überbrücken (z. B. in Teilkorpus 20: Spitteler *Mädchenfeinde* „Während dessen guckte ihr Gerold schelmisch in die Augen") oder Szenerien eröffnen (z. B. in Teilkorpus 20: Timm *Kopfjäger* „die See vor Augen").

Aber auch von den Sichtachsen abgesehen sind Referenzen auf Figuren und auch einzelne Körperteile derselben sehr häufig. „Hand", „ihm", „sich", „mir", „Kopf", „Gesicht" und „Herz" wurden sehr oft als Raumhinweise annotiert (Platz zwei bis sieben der häufigsten Annotationen dieser Kategorie). Hier zeigt sich, dass die Idee des Menschen als Ursprung und Zentrum der Raumwahrnehmung quantitativ für die Kategorie der Raumhinweise besonders wichtig ist. Dieser über Tuan (1979), Lotman (1973), Kant (2006 und 1993) und Husserl (1991) (vgl. Abschnitt 2.3) ins Modell eingebrachte Ansatz konnte also für die literarische Raumgestaltung als besonders fruchtbar ausgemacht werden.

Bereits die Netzwerkanalysen der Raumkategorien in den vier Teilkorpora haben gezeigt, dass Raumhinweise häufig in sehr engem Zusammenhang mit Relationen und relationalen Verben stehen. Betrachtet man nun den Kontext der in dieser Kategorie annotierten Wörter genauer, so wird deutlich, dass sie tatsächlich häufig als Knotenpunkte einer netzwerkartigen Raumstruktur fungieren. Insgesamt passen die Funde gut zu den Beobachtungen der Kategorie „relationale Verben". Die Sicht ist insgesamt bedeutender als das Rufen und Hören. Wie aber passt das häufige Vorkommen der „Hand" hier hinein? Das Tasten spielte unter den relationalen Verben schließlich eine sehr kleine Rolle („fühlen" kommt zwar deutlich häufiger vor, meint aber auch nur selten die taktile Wahrnehmung). Die Hände aber überbrücken häufig Zwischenräume und sind dadurch für Raumdarstellungen bedeutsam. Hände fahren über (eigene und fremde) andere Körperteile wie Gesichter, Lippen oder Augen (z. B. in Teilkorpus 20: May *Im Lande des Mahdi II* „legte sich, wie erschrocken, die Hand auf den Mund" oder Ganghofer *Ochsenkrieg*: „presste er die rechte Hand auf ihren Mund"). Sie greifen nach oder zwischen Dinge und werden anderen gereicht oder gegeben. Natürlich wird auch die ein oder andere Hand wieder zurückgezo-

gen. Darüber hinaus werden Hände zur Halterung von Objekten und Attributen der Figuren (z. B. in Teilkorpus 20: Wille *Abendburg* „Brennende Fackeln in der Hand“).

Bei der Betrachtung des Wortes „Hand“ lohnt sich aber auch der Blick in die Datenbank der Raummetaphern, die während des Trainingsprozesses angelegt wurde. Hier zeigt sich, dass räumliche Metaphern mit den Kernwörtern Hand und Hände relativ häufig sind. Im überhaupt sehr Raum-Metapher-freudigen 18. Jahrhundert kommen in elf von 20 Romanen aus dem Trainingskorpus 17 solcher Metaphern vor. Im 19. Jahrhundert gibt es weniger Raummetaphern mit diesen Kernwörtern (sieben Metaphern, davon sind zwei lediglich unterschiedliche Varianten einer einzigen Metapher) und diese kommen in weniger Texten vor (fünf von 20 Romanen aus dem 19. Jahrhundert im Trainingskorpus). In neun von 20 Romanen aus dem 20. Jahrhundert, die Teil des Trainingskorpus waren, kommen zehn unterschiedliche raummetaphorische Phrasen mit den Kernwörtern „Hand“ oder „Hände“ vor. Im 21. Jahrhundert gibt es schließlich nur noch eine Metapher dieser Kernworte in einem Roman. Es zeigt sich außerdem, dass Raummetaphern mit den Kernwörtern „Hand“ und „Hände“ sehr flexibel genutzt werden. Die Wörter „Hand“ und „Hände“ werden nicht als Ein-Wort-Metaphern genutzt, sondern sind Teil von variablen Phrasen. Wie Abbildung 8.3 zeigt, wird nur selten dieselbe metaphorische Phrase in Romanen unterschiedlicher Jahrhunderte verwendet (so der Fall bei „in jemandes Hand sein“ / „jemandem in die Hände fallen“, „alle Hände voll zu tun haben“ und „freie Hand haben“). Da das Wort „Hand“ in dieser Kategorie so häufig annotiert wurde, muss natürlich auch die Häufung der Verwendung in metaphorischen Kontexten berücksichtigt und die räumliche Bedeutung etwas relativiert werden.

Seltener als Hinweise auf Figuren oder einzelne Wahrnehmungsorgane sind im Kernkorpus solche, die Objekte referenzieren, wie z. B. Tische („Tisch“ 2.051 / „Tische“ 690), Stühle („Stuhl“ 720) oder Sofas („Sofa“ 330). Es wird sichtbar, dass die Idee von Raum als absoluter Größe, die mit Objekten angefüllt werden kann (Objektorientierter Raumbegriff wie in Abschnitt 2.3.1 beschrieben), hier weniger Platz findet als die netzwerkartige Struktur, die sich zwischen Figuren, deren Innen- und Außenwelten, und auch, wenn auch nachrangig, zwischen Figuren und Objekten oder zwischen Objekten ausbreitet. Bei der Analyse zeigt sich einerseits sehr klar, welche theoretischen Prämissen in die Erstellung des Classifiers eingeflossen sind, auf der anderen Seite zeigen die quantitativen Auswertungen wie bedeutsam genau diese Prämissen für literarische Texte sind, zeigen sie sich doch hochfrequent in der literarischen Darstellung von Raum. Es zeigt sich auch, dass Raumhinweise, die nicht unbedingt zu einem relationalen Raum beitragen, sondern z. B. ein landschaftliches Element enthalten (ein Beispiel wäre „Wellen“ mit 340 Annotationen), seltener sind als solche, die Teil eines relationalen Geflechts sind. Außerdem

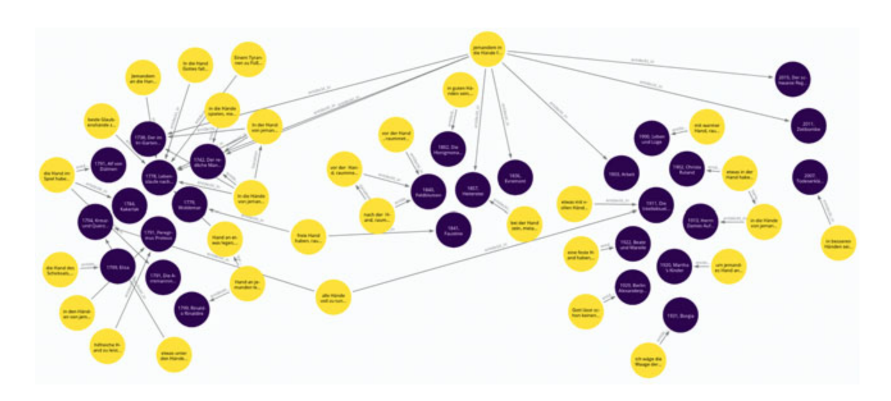


Abbildung 8.3 Raummetaphern mit den Kernwörtern „Hand“ und „Hände“ im Trainingskorpus

wird der Eindruck bestätigt, dass Rauminformationen oft implizit sind und durch Ausdrücke vermittelt werden, die neben diesen Rauminformationen noch andere Aspekte der Erzählung stützen. Der Satz „Er legte seine große Hand auf ihre kleine“ beinhaltet z. B. neben der Rauminformation darüber, wie viel Platz die Hände einnehmen und wie sie zueinander positioniert sind, noch eine emotionale Dimension, die dem Lesenden eröffnet, wie nahe sich die Figuren stehen und dass es hier wahrscheinlich um eine Situation geht, in der eine Figur größer und stärker ist und die kleinere, schwächere beschützen möchte. Wie im vorigen Abschnitt bereits ausführlich beschrieben, sind also auch Raumhinweise als wichtiger Teil der Darstellung des (relationalen) Raumes mehrdeutig. Dieser Aspekt wird auch bei der späteren Berechnung der RIWs berücksichtigt und die Bedeutung für die Raumdarstellung insgesamt dadurch etwas verringert. Die Erkenntnisse einer solchen übergreifenden Analyse schließen direkt an die konzeptionellen Überlegungen zum literarischen Raum an, die in Kapitel 2 betrachtet wurden. Der digitale Mixed-Methods-Ansatz stützt einige der raum- und literaturtheoretischen Thesen und spezifiziert, welches Raumverständnis literarischen Darstellungen zu Grunde liegt.

8.1.2 Innen- und Außenwelten

Bereits im vorangegangenen Kapitel wurde deutlich, dass die Körperteile von Figuren, die als Raumhinweise annotiert wurden, nicht nur Teil eines relationalen Raumgefüges sind. Sie sind auch wichtige Verortungspunkte für Innen- und Außenwelten

und deren permeable Grenzen. Häufig werden Geschehnisse im Inneren von Figuren oder in deren Umgebung verortet. Nun sind Innenwelten von Figuren zwar ein zentraler literarischer Aspekt, der maßgeblich zum Lesevergnügen beiträgt (Lahn und Meister, 2008, 232), sie sollen hier aber dennoch nicht ausführlich, sondern lediglich in ihrer räumlichen Verankerung, betrachtet werden. Das heißt, dass auch hier wieder konkrete Indikatoren auf Wortebene im Zentrum stehen. Ein wichtiger Faktor, der in der literaturwissenschaftlichen Forschung auch bereits als solcher ausgemacht wurde, ist die „prinzipielle Unzugänglichkeit der fremden Innenwelt" (Matzat, 2014, 20), die der Darstellung des Bewusstseins von Figuren einen besonderen Reiz verleihe (Matzat, 2014, 20–21). Gegenstandsbedingt liegt der Fokus narratologischer Ansätze zumeist auf der Darstellung figuraler Innenwelten, die mehr oder weniger durch die Erzählinstanz vermittelt werden können oder auf Innenwelten ebendieser Erzählinstanz. Im Zentrum steht häufig die Frage, wie viel die Erzählinstanz von einer oder mehreren Figuren weiß und wie die Erzählsituation gestaltet ist (vgl. Genette und Knop, 1998, 134ff und 232; Martínez, 2007, 64; Lahn und Meister, 2008, 126; Matzat, 2014, 21). Neben dem Wissen der Erzählinstanz ist auch das Sehen und das Sprechen der Figuren von zentraler Bedeutung (Schmid, 2014, 122). Zusammenfassend lässt sich also festhalten, dass sowohl Innen- als auch Außenwelten und vor allem die Frage, was, wie und wie viel einer Innenwelt nach außen dringt, von zentralem Interesse für die Erzähltextforschung ist. Eine detaillierte Untersuchung der Darstellung mentaler Ereignisse findet sich bei Schmid (2017). Schmid entwirft sechs Schablonen der Bewusstseinsdarstellung, die in zwei Gruppen – die markierte und die kaschierte Bewusstseinsdarstellung – fallen (Schmid, 2017, 55). Außerdem stellt Schmid fest, dass mentale Ereignisse auf Ebene der Exegese und auf Ebene der Diegese dargestellt werden können (Schmid, 2017, 81). Für diese Studie ist außerdem relevant, dass Schmid herausarbeitet, dass das Bewusstsein und damit eine Innenwelt auch als Ort oder aber als Motivierungsfaktor einer darzustellenden Zustandsveränderung fungieren kann (Schmid, 2017, 11). Mit der hier entwickelten Methodik kann weniger die genaue Erzählstruktur betrachtet werden. Stattdessen wird von der Einzeltextbetrachtung Abstand genommen, um die häufigsten Indikatoren der Innenwelt-Verortung in einem größeren Textkorpus anzuschauen. Ob diese Indikatoren Teil der Erzähler- oder der Figurenrede sind, ob hier mentale Prozesse mehr oder weniger vermittelt dargestellt werden, wird nicht analysiert. Stattdessen rückt eine auf die Figuren zentrierte Perspektive in den Vordergrund, die eher zum Ziel hat, eine verborgene Struktur im Text sichtbar zu machen, nämlich die relationale und dynamische Konstellation von Innen- und Außenwelten der Figuren, die auf unterschiedliche Weise für andere Figuren oder auch die Erzählinstanz verborgen oder zugänglich sind. An diese Betrachtungen anschließen könnte aber z. B. eine genauere Klassifizierung der Annotationsdaten,

die auf Innen- und Außenwelten referenzieren, im Hinblick darauf, ob sie Teil der Erzähler- oder Figurenrede sind und ob sie verbalisiert werden oder zur Gedankenrede gehören oder ob sie zur Darstellung von Handlungen gezählt werden können.

Literarische Außenwelten sind sehr divers und können unter unterschiedlichen Gesichtspunkten betrachtet werden. Zwei wichtige Faktoren von Außenwelten sind z. B. die Verortung von Figuren im Hier und Dort, die in Abschnitt 8.1.3 erläutert wird oder die Darstellung gesellschaftlicher und privater Räume, die in Abschnitt 8.1.6 genauer betrachtet wird. An dieser Stelle werden zu literarischen Außenwelten zunächst ein paar kleinere literaturwissenschaftliche Ansatzpunkte gewählt, um vor allem der Frage nachzugehen, inwiefern diese mit den Annotationsdaten dieser Studie zusammengeführt werden können und ob sich hier gute Einstiegspunkte in weiterführende Interpretationen bieten.

Innenwelten

Im Teilkorpus des 18. Jahrhunderts steht für die Innenwelt am häufigsten das Herz, gefolgt von Seele und Kopf.[9] Das Herz ist mit 2.982 Nennungen von „Herz" und 1.387 von „Herzen" aber mit deutlichem Abstand die meistgenannte Innenverortung; die beiden Wörter stehen in diesem Teilkorpus sogar auf Platz eins und Platz sieben der am häufigsten annotierten Raumhinweise insgesamt. Die Seele wurde im gesamten ersten Teilkorpus 1.135 Mal als Raumhinweis annotiert, der Kopf 920 Mal. Das Herz ist mit 1.146 Annotationen von „Herz" und 674 von „Herzen" auch im Teilkorpus des 19. Jahrhunderts wieder wichtigstes Verortungsorgan für das Innenleben. Der Kopf (1.044 Annotationen als „Raumhinweis") kann hingegen sowohl für eine Innenverortung („etwas geht durch den Kopf") stehen als auch über seine Beweglichkeit Innenwelten ausdrücken („den Kopf gesenkt halten") oder den Blick darauf verwehren („den Kopf abwenden"). Auch hier wird wieder offenbar, wie wichtig die Blickachsen zwischen den und auf die Figuren sind (vgl. Abschnitt 8.1.1). Die Seele nimmt hingegen mit 493 Annotationen im Teilkorpus des 19. Jahrhunderts deutlich an Bedeutung für die Innenverortung ab. Im diachronen Verlauf wird das Herz als wichtige Innenverortung immer seltener aufgerufen. War „Herz" im Teilkorpus des 19. Jahrhunderts noch auf Platz acht der am häufigsten annotierten Raumhinweise, so steigt es im Teilkorpus des 20. Jahrhunderts auf Platz 13 ab. Im letzten Teilkorpus findet es sich mit 299 Annotationen auf dem 35. Platz. Verzeichnet das Teilkorpus des 20. Jahrhunderts immerhin noch

[9] Als Indikatoren für Innen- und Außenwelten werden in diesem und im folgenden Abschnitt Ausdrücke betrachtet, die vom Raum-Classifier als „Raumhinweis" annotiert wurden. Eine tabellarische Übersicht über alle Annotationen dieser Kategorie ist im GitHub-Repository zu dieser Arbeit im Ordner „Datenbasis der Analyse / Raumhinweis" enthalten (Schumacher, 2021a).

638 Annotationen von „Herz“ und 337 von „Herzen“, so sind es im 21. Jahrhundert nur noch 299 Annotationen von „Herz“ und 108 von „Herzen“. Ein ähnliches Bild zeigt sich für „Seele“. Im ersten Teilkorpus mit 1.135 Annotationen noch auf Platz neun, rückt es mit 493 Annotationen im zweiten und 198 Annotationen im dritten und 53 Annotationen im vierten Teilkorpus weit in den Hintergrund.

Trotzdem sind sowohl „Herz“ als auch „Seele“ nach wie vor wichtige Hinweise auf Innenwelten der Figuren. Im Teilkorpus des 20. Jahrhunderts sind die Annotationen von „Seele“ z. B. auf 20 der 25 Romane verteilt und kommen damit in 80% der Texte des Korpus vor. Im Teilkorpus des 21. Jahrhunderts finden sich in 17 Romanen Annotationen von „Seele“, also in 68% der Romane. Auf der Außenseite bzw. an der Grenze zwischen Innenwelt und Außenwelt befindliche Körperteile wie Augen und Hände oder auch das Gesicht sind hier aber deutlich häufiger Teil der literarischen Raumhinweise.

Die Tatsache, dass Herz und Seele als Innenverortungen an quantitativer Bedeutung verlieren, könnte vermuten lassen, dass stattdessen der Kopf, der ebenfalls als solche dienen kann, in den Vordergrund rückt. Tatsächlich wurde das Wort „Kopf“ im ersten Teilkorpus nur 920 Mal annotiert, im zweiten dann 1.044 Mal, im dritten 1.103 und im vierten 1.679 Mal (vgl. Abbildung 8.4).

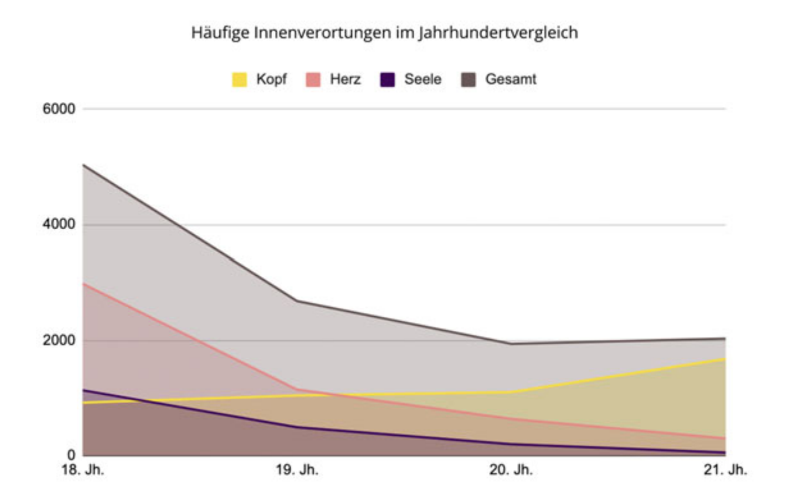


Abbildung 8.4 „Kopf“, „Herz“ und „Seele“ als Indikatoren der Innenverortung

Ausschließlich auf Grundlage dieser quantitativen Einblicke in die vier Teilkorpora ist es nicht möglich, eine genaue narratologische Analyse von Bewusstseinsdarstellungen im Besonderen oder auch von Innenwelten im Allgemeinen durchzuführen, die auch die unterschiedlichen Ebenen von Diegese und Exegese oder die Darstellungsformen mit einbezieht. Dennoch stützen die Daten interessanterweise literarhistorische Beobachtungen aus Schmids (2017) umfassender Betrachtung von Bewusstseinsdarstellungen. Es zeigt sich eine Zunahme des Ausdrucks „Kopf" als Raumhinweis in genau der Zeitspanne, die Schmid als Beginn einer Verlagerung der Handlung von der Außen- in die Innenwelt ausgemacht hat, nämlich im 19. Jahrhundert (Schmid, 2017, 1). Für das 18. Jahrhundert stellt Schmid eine „zunehmend sorgfältigere Motivierung der Handlung durch das Bewusstsein" (Schmid, 2017, 5) fest. Die in den vier Teilkorpora erhobenen Daten legen nahe, dass diese Handlungsmotivierung zuvor noch vergleichsweise häufig emotionaler Natur war. Je mehr das Bewusstsein zum Ort von Zustandsveränderungen wird, desto bedeutsamer wird der Kopf als Verortung mentaler Prozesse einerseits und als Raumhinweis andererseits. Im Gegenzug nimmt die Erwähnung der emotionaler konnotierten Ausdrücke „Herz" und „Seele" ab (auch wenn beide ebenfalls häufig bleiben). Anders als das Herz aber evoziert der Kopf nicht per se eine Innenverortung oder den Hinweis auf Emotionen, da er auch von außen betrachtet werden kann. So steht z. B. in dem Satz „Denken Sie sich auf dieser kleinen Gestalt einen, vielleicht im Verhältnis etwas zu großen Kopf, mit einer festen, an den Schläfen kahlen Stirn, unter der ein paar stechende Augen hevorblitzten [...]" (Teilkorpus 19: Spielhagen *Problematische Naturen. Zweite Abteilung*) der Kopf in einem das äußere einer Figur beschreibenden Kontext. Tatsächlich wurde das Wort „Kopf" im ersten Teilkorpus noch 136 Mal in Kombination mit „in" oder „im" im unmittelbar davor stehenden Kontext genutzt. Im zweiten Korpus waren es noch 42 Mal und im dritten nur noch 37 Mal. Im vierten waren es dann schließlich wieder 131 Mal. Da aber Bewusstseinsdarstellungen keine explizite Innenverortung benötigen, sondern auch unmarkiert bleiben können (Schmid, 2017, 55), können die quantitativen Daten dieser Studie diesbezüglich auch nicht als belastbares Fundament für eine Interpretation dienen. Wieder kann hier nur ein Einstieg in eine solche gezeigt werden. Die mit Hilfe des Raum-Classifiers erhobenen Daten können zu möglichen Indikatoren literarhistorischer Phänomene werden und Hinweise auf einen quantitativen Abgleich von Hypothesen geben.

In allen vier Korpora sind Gesicht und Augen besonders häufig Grenzen der Innenwelt und gleichzeitig können sie die Geschehnisse im Innern nach außen zeigen oder äußere Eindrücke hineinlassen. Wie im vorangegangenen Kapitel bereits beschrieben, werden Augen in zweifacher Hinsicht raumsymbolisch eingesetzt. Erstens werden Blickachsen ausgedrückt (z. B. Teilkorpus 18: Wieland *Geschichte des*

Agathon „mit ihren Augen gesehen haben“, Huber *Die Familie Seldorf* „wo ihr Auge umherspähte“) und zweitens Innen- und Außenwelt verbunden, bzw. Innenwelten dargestellt (z. B. Teilkorpus 20: Ganghofer *Ochsenkrieg* „die Amtmännin machte erstaunte Augen“ oder Tucholsky *Schloß Gripsholm* „der allgemeine Irrsinn in den Augen“). Dazu gehört auch, dass eine solche Verbindung abgeschnitten werden kann (Teilkorpus 18: Moritz *Andreas Hartknopf* „schlugen beschämt ihre Augen nieder“, Miller *Siegwart* „und schlug schnell das Auge nieder“). Das Gesicht verbindet auf ähnliche Weise Innen- und Außenwelt (Teilkorpus 18: Wieland *Geschichte des Agathon* „ihr Gesicht war traurig“, ebd. „als sie schon in seinem Gesicht etwas bemerkte“, Schiller *Der Geisterseher* „Ernst, der das ganze Gesicht überschattete“), muss aber zur Eingrenzung der Innenwelt aktiv verdeckt werden (Teilkorpus 19: Eichendorff *Ahnung und Gegenwart* „und bedeckte ihr Gesicht mit beiden Händen“, Teilkorpus 18: Miller *Siegwart* „hielt den Hut halb vors Gesicht“).

Weniger häufig als Augen und Gesicht oder Herz, Seele und Kopf geben die Hände Hinweise auf Innenwelten. Zittern diese jedoch, so meist im Kontext eines nicht mehr zu verbergenden inneren Geschehens (z. B. Teilkorpus 18: Millers *Siegwart* „Meine Hand zitterte vor Erwartung“). Raumnarratologisch interessant ist hier, dass zwei von Lotman beschriebene Raum-Motive zusammenwirken (vgl. Abschnitt 2.2.1). Einerseits steht das wahrnehmende Individuum im Zentrum, auf der anderen Seite erstreckt sich die Grenzthematik nicht nur auf den Außenraum, sondern kann auch im Zusammenspiel von Innen- und Außenraum erkannt werden. Die Grenzen sind hier in erster Linie die Augen, in zweiter das gesamte Gesicht. Weniger bedeutsam, aber auch vorkommend, sind weitere Wahrnehmungsorgane (wie z. B. die Hände). Ergänzend zu Lotmans achsenbasiertem Raummodell kann festgestellt werden, dass die Verortung im Innenraum nicht mit positiv-negativ-Polen verbunden ist. Stattdessen werden Herz und Seele meist als emotionale Zentren dargestellt, der Kopf hingegen als Sitz des Rationalen. Natürlich kann aber beides auch als Verkehrung ins Gegenteil dargestellt werden (Teilkorpus 18: Gellert *Das Leben der schwedischen Gräfin*, Teilkorpus 19: de la Motte Fouqué *Die Frau des Falkensteins* und Marlitt *Goldelse* „kaltes Herz“ und Teilkorpus 18: Jean Paul *Hesperus* und *Blumen, Frucht- und Dornenstücke* „hitziger Kopf“).

Die Verortung von Innenwelten kann mit Hilfe der automatischen Annotation durch den Raum-Classifier in Form von Indikatoren auf der Wortebene identifiziert werden. Dadurch wird eine vergleichende Analyse des Wortgebrauchs möglich. Auch Rückschlüsse über die Verbindungen zwischen den Innenwelten unterschiedlicher Figuren können auf dieser Basis gezogen werden. Wieder spielen die bereits häufig erwähnten Sichtachsen eine besondere Rolle, da sie Relationen aufbauen und in das Innere einer Figur eindringen können. Genauere Charakterisierungen der Innenwelten werden erst möglich, wenn die Distanz zum Einzeltext verringert

wird und der Fokus auf den spezifischen Kontext der Indikatoren von Innenwelten gelegt wird. Die bereits genannte Klassifizierung der Indikatoren nach narratologischen Kategorien wäre ein möglicher Zwischenschritt, der auch mit einer mittleren Distanz zum Korpus und einer etwas geringeren Anzahl an Texten insgesamt durchführbar wäre. Eine andere Möglichkeit der Verknüpfung des hier entwickelten Distant-Reading-Ansatzes mit bisheriger literaturwissenschaftlicher Forschung über figurale Innenwelten wäre, mit Hilfe des Raum-Classifiers, eine kleine Anzahl von Texten auszumachen, in denen besonders viele Innenwelt-Indikatoren vorkommen und diese dann im Close-Reading-Verfahren auf die genaue Ausgestaltung innerer Vorgänge und die diesbezügliche Verknüpfung der Figuren untereinander zu untersuchen.

Außenwelten (Dingwelten)
Bisher wurden ausschließlich Phänomene betrachtet, die durch quantitative Aspekte auffallen. Dieser Ansatz ist sehr gängig in den digitalen Literaturwissenschaften (erkennbar z. B. in den Studien von Moretti (2013), Jockers (2013), Underwood (2019) oder Piper (2018)), in nicht digitalen Untersuchungen finden sich aber eher Erkenntnisse, die sich auf Betrachtungen ungewöhnlicher Ausdrücke und Passagen stützen. Für den literarischen Raum des 19. Jahrhunderts haben z. B. Seeburg und Jürjens (beide 2018) postuliert, dass Dingwelten von Bedeutung seien. Dingwelten werden durch besondere Spuren individuellen Lebens konstruiert (vgl. Seeburg, 2018, 46; Jürjens, 2018, 36–37). Nun finden sich z. B. im Teilkorpus des 19. Jahrhunderts auch Gegenstände unter den am häufigsten als Raumhinweis annotierten Wörtern. Dazu gehört mit 587 Annotationen auf Platz 21 der Tisch. Das Bett wurde in diesem Teilkorpus nur 273 Mal annotiert, der Stuhl nur 239 Mal. Aber für die Dingwelten sind diese häufig vorkommenden Objekte auch weniger von Bedeutung als weniger häufig vorkommende, individuellere Ausdrücke. Wörter wie „Nähtischchen", „Tabaksdose", „Kaminverzierungen", „Weihgefäßen", „Marmorfliesen", „Billardtische", „Pistolenkästchen", „Krückenstock" oder „Samtröckchen" wurden jeweils nur ein einziges Mal im gesamten Teilkorpus annotiert, tragen aber eher zu besonderen Spuren und damit zu Dingwelten bei (vgl. Seeburg, 2018, 46; Jürjens, 2018, 36–37). Nun gibt es unter den annotierten Wörtern sehr viele, die nur ein einziges Mal annotiert wurden. Im Teilkorpus des 19. Jahrhunderts, das hier herangezogen wird, weil es dem Forschungsgegenstand entspricht, an dem die Dingwelten-These entwickelt wurde, sind es 8.750 Wörter, die eine Annotationsfrequenz von 1 im gesamten Teilkorpus aufweisen; genug einzigartige Wörter, um Dingwelten in 25 Romanen auszugestalten. Zwar legt auch die hier durchgeführte Studie einen Schwerpunkt auf quantitativ bedeutsame Phänomene, es soll dabei aber nicht vergessen werden, dass die Summe der Hapax Legomena ebenfalls

eine beachtliche Größe erreichen kann. Eine These wie die Dingwelten-These kann durch die Betrachtung dieser selten annotierten Raumhinweise gestützt werden. Mit Hilfe des Raum-Classifiers könnte aus den einmalig annotierten Raumhinweisen, die zu Dingwelten beitragen, auch eine Liste erstellt werden, die dann für einen weiterführenden listenbasierten Ansatz genutzt werden kann. Eine andere Möglichkeit wäre, die Distanz zum Korpus zu verringern, um die Dingwelten-These genauer zu betrachten. Es könnte mit dem Raum-Classifier zum Beispiel ein Text ausgemacht werden, in dem besonders viele Raumhinweise nur ein einziges Mal annotiert wurden. Dieser Einzeltext könnte dann genauer betrachtet werden. Eine solche Vorgehensweise kann aber auch genutzt werden, um zu überprüfen, ob eine These, die anhand eines Einzeltextes oder eines sehr kleinen Korpus entwickelt wurde, überhaupt auf andere Texte übertragbar ist. In diesen Beispielen zeigt sich auch das heuristische Potential digitaler Textanalysen wie es z.B von Kuhn (2018), Gius (2020) und Gerstorfer (2020) beschrieben wird.

Eine These wie die, dass im 19. Jahrhundert „vor allem das weiße Leinen und die damit verbundenen Prozesse von Reinigung und Bleiche stapelweise Eingang in die Literatur“ (Jürjens, 2018, 38) finden, hält einer Korpusanalyse z. B. nicht stand. Leinen und Leinenartikel wurden bei einer von den automatisch durchgeführten Annotationen unabhängigen Abfrage einer kurzen Liste von Wörtern in CATMA, die „Leinen“ und „Baumwolle“ inkludiert, im gesamten Teilkorpus des 19. Jahrhunderts lediglich 24 Mal in insgesamt zehn der 25 Romane gefunden. Baumwolle und Baumwollprodukte 31 Mal in fünf Romanen, „reinigen“ lediglich 29 Mal in 9 Romanen. Gebleichtes Leinen kommt nur in seiner Negativform, also als ungebleichtes Leinen, vor. Zwar gibt es in der Tat eine Häufung der genannten Begriffe in Kellers *Sinngedicht*, welches Jürjens zum Ausgangspunkt ihrer Analyse macht, allerdings zeigt sich hier die begrenzte Übertragbarkeit solcher Thesen auf eine relativ große zeitliche Spanne literarischen Schaffens. Die hier später noch genauer dokumentierte Analyse der Raumthemen zeigt, dass auch die Thematik der Ordnung in diesem Teilkorpus nicht häufiger aufgerufen wird als im Teilkorpus des 18. Jahrhunderts, sondern dass das Thema „Ordnung“ in der Literatur eher im diachronen Verlauf seltener aufgerufen wird (für das 18. Jahrhundert wurden hier 173 Annotationen zum Thema „Ordnung“ subsummiert, für das 19. Jahrhundert 104, für das 20. Jahrhundert 51 und für das 21. Jahrhundert 61 (vgl. Abbildung 8.11 der Thementafeln in Abschnitt 8.1.5 und das GitHub-Repository mit den Forschungsdaten zu dieser Arbeit). Auch hier zeigt sich also, dass Einzelfallanalysen nicht immer ohne Abgleich mit einem größeren Korpus auf größere Zusammenhänge übertragen werden können. Die Ordnung als zentrales Paradigma der Literatur des 19. Jahrhunderts auszurufen (Jürjens, 2018, 39), scheint in Anbetracht dieser Datenlage etwas gewagt. Eine Zusammenführung quantitativer Methodik im Allgemeinen

und in diesem speziellen Fall die Nutzung eines Tools wie des Raum-Classifiers mit Close-Reading-Analysen, ist dabei sehr vielversprechend.

8.1.3 Das Spannungsfeld zwischen dem Hier und dem Dort

Die Kategorie der Orte ist mit 175.182 Annotationen im gesamten Kernkorpus die vierthäufigste, fällt aber dadurch, dass sie Raum sehr explizit darstellt, auch qualitativ besonders ins Gewicht. Von überragender quantitativer Bedeutung ist in dieser Kategorie das Spannungsfeld zwischen dem Hier und dem Dort.[10] Zwar sind „hier" (10.342 Annotationen) und „da" (7.151 Annotationen) die beiden häufigsten der insgesamt 17.859 Types, die in dieser Kategorie annotiert wurden, es gibt aber noch weit mehr Wörter und auch kontextuelle Aspekte, die zu diesem Spannungsfeld beitragen. Es geht dabei auch um Darstellungen von Nähe und Ferne und darum, ob eher klar umrissene Raumeinheiten benannt werden oder vage, offene Räume.

Im Teilkorpus des 18. Jahrhunderts sind die Wörter „Hier" und „hier" (zusammen 2.464 Annotationen) häufiger als „Da", „da", „Dort" und „dort" (zusammen 1.901 Annotationen). Im Teilkorpus des 19. Jahrhundert dreht sich das Verhältnis um und das Dort (insgesamt 4.138 Annotationen) wird häufiger erwähnt als das Hier (insgesamt 3.996 Annotationen). Auch in den Teilkorpora des 20. und 21. Jahrhunderts sind die unterschiedlichen Ausdrücke, die das Dort referenzieren, häufiger, allerdings gibt es nur im Teilkorpus des 20. Jahrhunderts einen deutlichen quantitativen Unterschied zwischen Hier (2.598 Annotationen) und Dort (3.680 Annotationen). Dass im Teilkorpus des 18. Jahrhunderts „Hier" häufiger annotiert wurde, wird durch die Tatsache aufgewogen, dass mit „Welt" (1.941 Annotationen) ein Ausdruck am zweithäufigsten annotiert wurde, der ebenfalls als Gegenbegriff zu „Hier" genutzt werden kann, der aber sehr mehrdeutig ist (vgl. Abschnitt 8.1.4). Literarisch bedeutsam scheint vor allem das Spannungsfeld zwischen einem Hier und einem Dort, einer Nähe und einer Weite, zu sein. Beide Extreme dieses Gegensatzpaares sind – wie die Ähnlichkeit in der Häufung zeigt – nahezu in der Waage. Auch andere häufig als Ort annotierte Wörter tragen zu diesem Spannungsfeld bei. Erstaunlich ist dabei das korpusübergreifend relativ homogene Bild der häufigsten Ausdrücke.

Mit „Hause", „Haus", „Zimmer", „Stadt", „Weg", „Erde" und „Boden" stehen sich bei den nächstplatzierten der am häufigsten mit dem Ort-Tag versehenen Wörter im Kernkorpus Innen- und Außenräume gegenüber (vgl. Tabelle 8.3). Die Platzie-

[10] Eine Übersicht über alle mit dieser Kategorie annotierten Wörter und ihre Häufigkeiten in den einzelnen Teilkorpora und im gesamten Kernkorpus befindet sich im GitHub-Repository zu dieser Arbeit unter „Datenbasis der Analyse / Ort" (Schumacher, 2021a).

Tabelle 8.3 Am häufigsten annotierte Wörter in der Kategorie „Ort“ in den vier Teilkorpora

Rang	18. Jh	19. Jh	20. Jh	21. Jh
1	hier	hier	hier	hier
2	Welt	da	da	da
3	Himmel	dort	dort	Haus
4	da	Hause	Welt	Welt
5	Hause	Welt	Haus	dort
6	Zimmer	Haus	Hause	Hause
7	Erde	Zimmer	Weg	Zimmer
8	Weg	Stadt	Himmel	Weg
9	Haus	Weg	Zimmer	Stadt
10	Stadt	Boden	Boden	Boden

rungen in der Rangliste und die genaue Anzahl der Annotationen variieren zwar in den Teilkorpora, aber die Abweichungen sind erstaunlich gering. Die Verankerung im klar Umrissenen, Kleinen, Lokalen wird häufig der gesamten, großen, freien Welt gegenübergestellt. Dass dieser Fund auch Thesen aus den analogen Literaturwissenschaften entspricht, wird im folgenden Kapitel thematisiert. Da das Kernkorpus in dieser Hinsicht so homogen ist, wäre es für Anschlussuntersuchungen außerdem interessant, literarische Korpora aus anderen Kulturkreisen zum Vergleich heranzuziehen. Auf diese Weise könnte die Hypothese überprüft werden, dass es sich möglicherweise um ein kulturspezifisches Phänomen handelt.

Schauplätze in Form von einzelnen benannten Städten kommen zwar häufig vor, aber viel seltener als die zuvor genannten Ausdrücke (vgl. Tabelle 8.4). Welche Ortsnamen in den einzelnen Teilkorpora am häufigsten referenziert werden, ist zwar durchaus interessant, sagt aber bei einem Korpus von jeweils nur 25 Texten nicht viel aus. Schon die gehäufte Erwähnung eines einzelnen Ortes in einem einzigen Text kann reichen, um für diesen Ort Werte zu erzielen, die so hoch sind, dass sie unter den häufigsten Annotationen dieser Kategorie platziert werden. Dadurch lassen sich aber keine haltbaren allgemeinen Aussagen für das ganze Korpus ableiten. So schafft es z. B. Peking auf Platz drei des Teilkorpus des 20. Jahrhunderts, weil zwei Romane von Heyking ein entsprechendes Setting haben (vgl. dazu Tabelle 8.4 und auch Abschnitt 8.3).

Im Hinblick auf die Anwendung der Methode der Named Entity Recognition in der Literaturwissenschaft kann hier ein erster Hinweis darauf entdeckt werden, dass benannte Entitäten für die literarische Darstellung von Orten in Romanen

Tabelle 8.4 Häufigste drei Ortsnamen in den Teilkorpora

Teilkorpus	Platz 1	Platz 2	Platz 3
18. Jahrhundert	Paris 176	Berlin 142	London 103
19. Jahrhundert	Paris 332	Wien 185	Berlin 145
20. Jahrhundert	Hamburg 163	Berlin 109	Peking 62
21. Jahrhundert	Berlin 138	Paris 123	London 98

weniger bedeutsam sind als andere Raumkategorien. Statt konkret benannter Städte oder Länder oder auch Straßennamen werden stärker generische Benennungen von Orten wie „Natur", „Gegend" oder auch das nicht näher spezifizierte „Dorf" oder eben Standort-Bezeichnungen wie „hier" und „da" bevorzugt. Domänenadaptionen wie die hier vorgenommene sind also unverzichtbar, wenn computerlinguistische Methoden auf die Literaturwissenschaft übertragen werden.

8.1.4 Die abnehmende Bedeutung von Welt und Himmel für den literarischen Raum

In literaturwissenschaftlichen Analysen werden eher selten einzelne Begriffe genauer untersucht. Im Zusammenhang mit bedeutungshaltigeren Konzepten, können sie aber durchaus eine wichtige Rolle spielen, wie z. B. der Weltbegriff in Krobbs *Erkundungen im Überseeischen. Wilhelm Raabe und die Füllung der Welt*. Die vom Raum-Classifier unterstützte Analyse lenkt den Blick zuweilen recht stark auf einzelne Wörter, die auffallend häufig oder die in den Teilkorpora sehr unterschiedlich häufig annotiert wurden. Dass eine Einzelwortanalyse vor allem ergänzend zu Ansätzen, die ein ganzes Konzept hinter dem Wort in den Blick nehmen, wie die von Krobb, sinnvoll sein können, wird in diesem Teilkapitel anhand von zwei Fallstudien gezeigt. Neben Begriffen, die für bedeutungshaltigere Konzepte stehen, bieten sich auch Begriffe, die häufig metaphorisch gebraucht werden oder Kernworte von Metaphern sind, für eine solche Einzelwortanalyse an. In diesem Fall können die Annotationen des Raum-Classifiers mit einer Auswertung der Daten in der parallel zum Trainingsprozess aufgebauten Metaphern-Datenbank *laRa* zusammengeführt werden.

Im Teilkorpus des 18. Jahrhunderts ist das Wort „Welt" das am zweithäufigsten als „Ort" annotierte Wort (vgl. Tabelle 8.3 im vorigen Kapitel). Im diachronen Vergleich fällt auf, dass die quantitative Bedeutung des Wortes im Laufe der Jahr-

hunderte abnimmt. Ähnliches ist beim Wort „Himmel" zu beobachten. Beide Wörter werden häufig metaphorisch genutzt. Der Himmel ist darüber hinaus auch ein religiöser Begriff. Beim Himmel kommt hinzu, dass vor allem im literarhistorischen Kontext nicht immer ganz klar ist, ob der religiöse Begriff „Himmel" einen (jenseitigen) Ort bezeichnet oder eher eine Utopie, also einen nicht-Ort (Foucault, 2005, 11). Die Wandlung von der Bezeichnung eines Ortes, den Gläubige als klar verort- und betretbar klassifizieren würden, zu einer Metapher für einen seligen Zustand, muss hier in die Analyse einbezogen werden. Schon ein kurzer Blick auf eine Visualisierung sämtlicher Raummetaphern in *laRa* zeigt, dass sowohl „Himmel" als auch „Welt" (ebenso wie „hoch", „groß" und „Weg") zu den zentralsten Metaphern im Trainingskorpus gehören (vgl. Abbildung 8.5)

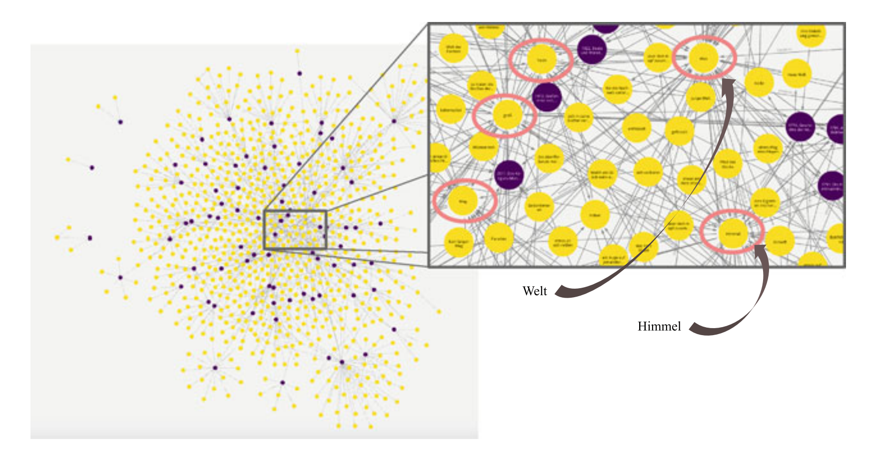


Abbildung 8.5 Visualisierung aller Metaphern in laRa mit Fokus auf das Zentrum

Zentral sind diese Metaphern vor allem, da sie mit vielen Variationen verknüpft sind, aber auch, da sie in vergleichsweise vielen Texten vorkommen. Bei den folgenden Fallstudien zu den Begriffskontexten wird darum auch genauer betrachtet, inwiefern der metaphorische Variantenreichtum und die vergleichsweise häufige metaphorische Verwendung in literarischen Texten bei einer digitalen literaturwissenschaftlichen Analyse Berücksichtigung finden kann.

Fallstudie 1: Weltbegriff bei Wilhelm Raabe und im gesamten Teilkorpus
Eine große Stärke quantitativer Ansätze ist es, dass kontrastive Analysen gemacht werden können, die die Untersuchung größerer Korpora und damit auch größerer

Datenmengen erlauben. Analoge Ansätze können dadurch sinnvoll ergänzt werden. Ein Beispiel für einen literaturwissenschaftlichen Ansatz, in dem der Weltbegriff eine große Rolle spielt, ist Krobbs *Erkundungen im Überseeischen*. Hierin wird die These aufgestellt, dass durch die beginnende Globalisierung im 18. und 19. Jahrhundert das Weltempfinden gleichzeitig sehr groß und sehr klein wurde; ferne Länder wurden zunehmend erfahrbar und fanden dadurch Einzug ins Lokale (Krobb, 2009, 9–10). Als Höhepunkt dieses kulturell gewandelten Weltempfindens wird das Ende des 19. Jahrhunderts definiert (Krobb, 2009, 10). Krobb erkennt darin ein wichtiges Motiv für die Literatur dieser Zeit (Krobb, 2009, 10). Als kulturellen Hintergrund nennt Krobb das „zweite Entdeckungszeitalter", dessen Beginn er für die Mitte des 18. Jahrhunderts festlegt (Krobb, 2009, 10). Zeitlicher Fokus ist also Mitte des 18. bis Ende des 19. Jahrhunderts. Raabes Werke fallen ins letzte Drittel dieses zeitlichen Rahmens. Um Raabes Werk auch literaturgeschichtlich einzuordnen und aufzuzeigen, dass auch andere Werke dieser Zeit den Weltbegriff und das Zusammenspiel von Nähe und Ferne thematisieren, werden bei Krobb Freytags *Soll und Haben* und Fontanes *Stechlin* genannt. Ähnlich wie in der Dialektik der gleichzeitig groß und klein empfundenen Welt wird hier also sowohl ein relativ großer zeitlicher Rahmen (etwa 150 Jahre) und eine relativ kleine Auswahl an ergänzenden Beispielen (zwei Romane) betrachtet. Die hier angewendete, quantitativ ausgerichtete Methode bietet die Möglichkeit, dieses Ungleichgewicht etwas auszugleichen. Um zunächst den literarhistorischen Kontext und die Verwendung des Weltbegriffs im diachronen Verlauf zu beleuchten, wird eine Einzelwortanalyse von „Welt" im Kernkorpus durchgeführt. Zusätzlich wurden für diese Fallstudie sechs Romane Wilhelm Raabes mit Hilfe des Raum-Classifiers annotiert und in Bezug zum Kernkorpus gesetzt.

Mit 1.941 Annotationen wurde „Welt" im Teilkorpus des 18. Jahrhunderts nach „hier" am zweithäufigsten annotiert (vgl. Tabelle 8.3). Im Teilkorpus des 19. Jahrhunderts steht der Begriff mit 1.378 Annotationen auf Platz fünf der Kategorie Ort, auf Platz vier steht er mit 804 Annotationen im Teilkorpus des 20. Jahrhunderts und im 21. Jahrhundert ist er mit 1.071 Annotationen ebenfalls auf Platz vier. Im gesamten Teilkorpus ist „Welt" also ein wichtiger Begriff bzw. quantitativ bedeutsam für die Raumkategorie „Ort". Dennoch fällt die Häufigkeit vom 18. zum 20. Jahrhundert ziemlich stark ab und steigt dann wieder etwas an. Dieser Trend bleibt auch dann bestehen, wenn die Anzahl der Nutzung des Begriffes „Welt" im gesamten Kernkorpus pro Roman und in Relation zu dessen Gesamtlänge betrachtet wird (vgl. Abbildung 8.6) Es lohnt sich aber, zunächst die Kontexte etwas genauer anzusehen.

Werden die 200 am häufigsten mit dem Tag „Ort" versehenen Wörter im Teilkorpus des 18. Jahrhunderts in Form von Small Multiple Wordclouds visualisiert, so zeigt sich auf den ersten Blick, dass „Welt" tatsächlich in der Mehrzahl der Texte zu den am häufigsten genannten Orten gehört (vgl. Abbildungen 30 im Vergleich zu

31 und 33 im elektronischen Zusatzmaterial). Im Teilkorpus des 19. Jahrhunderts wurde „Welt“ sehr viel seltener als Ort annotiert. Neben dem häufigsten Wort „hier“ stehen davor in der Rangliste der am häufigsten annotierten Orte noch „da“, „dort“ und „Hause“. Mit „Haus“, „Zimmer“ und „Stadt“ stehen auch auf den folgenden Plätzen Wörter, die ein heimatliches Umfeld beschreiben können (vgl. Tabelle 8.3). Die Daten bestätigen damit Interpretationshypothesen wie die von Krobb, dass die literarische Welt im 19. Jahrhundert eher im Lokalen, Kleinen, Heimatlichen gezeigt wird (Krobb, 2009, 35). Die oben bereits erwähnte These, dass im 19. Jahrhundert ein Höhepunkt der Verwendung des literarischen Weltmotivs zu erkennen ist (s.o.), kann dadurch aber nicht unterfüttert werden. In den Teilkorpora 20 und 21 sieht es sehr ähnlich aus. Begriffe wie „hier“, „da“, „dort“, „Welt“, „Haus“ und „Hause“ stehen in fast gleicher Reihenfolge auf den ersten Plätzen der Rangliste von Ortsannotationen (vgl. Tabelle 8.3). Immer scheint also der Gegensatz zwischen häuslicher, kleiner, lokaler Welt und großer, globaler Welt für den literarischen Raum konstitutiv zu sein. Für das 18. Jahrhundert wurde bereits gezeigt, dass der Begriff „Welt“ in vielen der Romane des Teilkorpus von überragender Bedeutung ist. Neben den hohen Vorkommnissen in einigen oder sogar vielen Romanen gibt es aber auch eine Breitendeckung, die sich nicht auf das 18. Jahrhundert beschränkt. Die Analyse des Begriffes „Welt“ pro Roman hat gezeigt, dass in allen Romanen des Kernkorpus mindestens ein Mal das Wort „Welt“ als „Ort“ annotiert wurde. Es nimmt also nicht die Anzahl der Erzähltexte ab, in denen „Welt“ referenziert wird, aber die Häufigkeit mit der das geschieht. Folgt man in der Interpretation der Daten dem Ansatz von Krobb, so lässt sich argumentieren, dass mit Beginn der Globalisierung und dem „zweiten Entdeckungszeitalter“ der Begriff „Welt“ geradezu inflationär in der Literatur verwendet wird. Ab dem 19. Jahrhundert setzt dann die von Krobb beschriebene Dialektik von Welt im Lokalen, von Nähe und Ferne ein. Im gesamten 20. Jahrhundert bleibt die Bedeutung des Begriffes für den literarischen Raum ungefähr gleich, bis sie im 21. Jahrhundert, einer durch und durch globalisierten Zeit, wieder etwas ansteigt.

Nach all diesen Einblicken in die quantitative Verteilung des Weltbegriffs im Korpus, ist nun eine vergleichende Einordnung von Wilhelm Raabes Romanen möglich. Dazu wurden sechs Romane des Autors, die nicht zum Kernkorpus gehören, zum Teil aber Teil des Trainingskorpus waren, mit Hilfe des Raum-Classifiers annotiert und in CATMA hochgeladen. Bei dem Text, der Teil des Trainingskorpus war, handelt sich um *Die Akten des Vogelsangs* (vgl. Auflistung der Trainingsdaten in Tabelle 1 des elektronischen Zusatzmaterials). Für den im Trainingskorpus enthaltenen Erzähltext ist es wahrscheinlich, dass die Erkennung durch den Classifier bessere Resultate erzielt als bei den anderen Texten. Da *Die Akten des Vogelsangs*

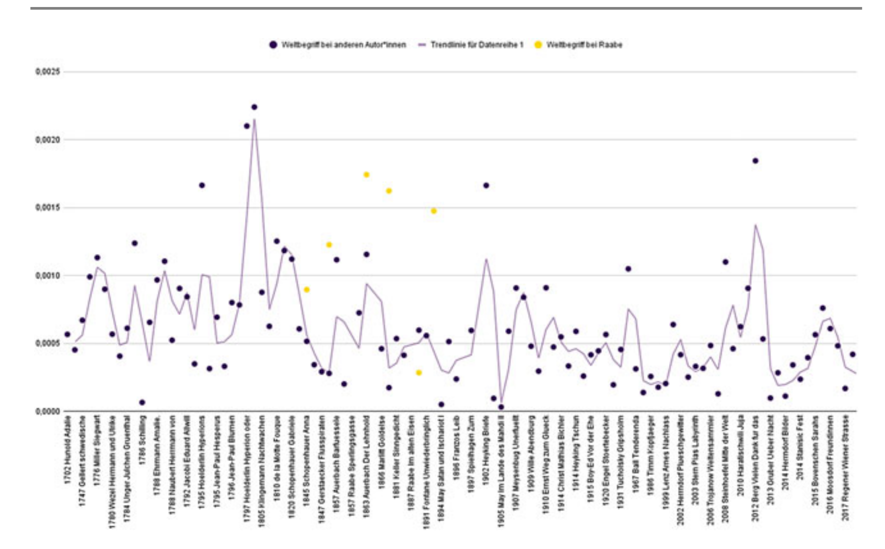


Abbildung 8.6 Übersicht der Annotationen des Wortes „Welt“ im Kernkorpus und in sechs Raabe-Romanen

aber der Text mit den niedrigsten Annotationen des Begriffes „Welt“ ist, fällt das in dieser Fallstudie nicht ins Gewicht.

Wie im Teilkorpus des 18. Jahrhunderts steht auch in diesem kleinen Raabe-Korpus der Weltbegriff an zweiter Stelle der Annotationshäufigkeiten der Kategorie Ort. Der Begriff wurde in unterschiedlicher Häufigkeit in jedem der sechs Raabe-Romane annotiert. Die absoluten Annotationszahlen wurden normalisiert, indem sie durch die Gesamtzahl der Worte im Text geteilt wurden. Diese Werte wurden dann in eine Übersicht des ebenso aufbereiteten, ganzen Kernkorpus eingetragen, sodass sich das Bild in Abbildung 8.6 ergab: Die Trendlinie, die an den als dunkle Punkte dargestellten Welt-Annotationen im Kernkorpus ausgerichtet ist, zeigt einen gleitenden Durchschnitt und macht den Abfall vom 18. zum 19. Jahrhundert gut sichtbar. Die hellen Punkte, die für Raabes Romane stehen, zeigen im Vergleich nur zum Teil ein typisches Verhalten für ihren zeitlichen Wirkungskontext. Vier der sechs annotierten Raabe-Romane weisen eine deutlich höhere Verwendung des Weltbegriffs auf als andere Romane der zweiten Hälfte des 19. Jahrhunderts. Nur zwei der Raabe-Romane lassen sich in den generellen Trend einordnen. Zwar sind Raabes Romane in der Tat ein sehr gutes Beispiel, um sich der Welt als literarischem Motiv zu nähern, wie Krobb das in seiner Analyse zu Erkundungen im Überseeischen tut, allerdings scheint der literarhistorische Höhepunkt der Auseinandersetzung mit der Welt nicht

in der zweiten Hälfte des 19., sondern rund 100 Jahre früher, in der zweiten Hälfte des 18. Jahrhunderts, zu liegen. Zwar erreichen Raabes Romane nicht die Ausreißer-Positionen von Romanen des 18. Jahrhunderts wie Hölderlins *Hyperion*-Texte und Novalis' *Die Lehrlinge zu Sais*, sie bilden aber zusammen mit Auerbachs *Barfuessele* und *Lehnhold* sowie Heykings *Briefe, die ihn nicht erreichten* eine Gruppe von Erzähltexten, in denen die Welt vergleichsweise häufig referenziert wird. Texte in denen „Welt" ähnlich häufig wie „Ort" annotiert wurde, finden sich dann erst wieder im 21. Jahrhundert mit Steinhöfels *Mitte der Welt* (2008) und vor allem Bergs *Vielen Dank für das Leben* (2012).[11] Für eine abschließende Aussage dazu ist allerdings auch die Stichprobe des Kernkorpus dieser Untersuchung zu klein. Denkbar wäre ein Vergleich mit mehreren 100 oder gar 1000 Romanen. Doch auch in diesem kleinen, beispielhaften Rahmen zeigt sich schon, wie sinnvoll ein solcher Abgleich sein kann, um nicht allzu schnell vom Besonderen zum Allgemeinen zu schließen und um Text- und Autor*innen-übergreifende Thesen zu untermauern. Ein großer Unterschied zwischen den Auswertungen der Annotationsdaten und einer hypothesengeleiteten Analyse wie der von Krobb ist, dass hier alle als Ort annotierten Vorkommnisse des Wortes „Welt" betrachtet werden und nicht nur diejenigen, die in einem Globalisierungs- oder Kolonialkontext stehen. Die Häufungen im 18. Jahrhundert könnten z. B. dadurch relativiert werden, dass der Begriff hier in grundsätzlich anderen Kontexten gebraucht wird, in denen damit nicht wirklich die gesamte Welt gemeint ist. Oder es könnte sein, dass „Welt" besonders häufig metaphorisch genutzt wird. Beide Möglichkeiten sollen nun noch kurz in Betracht gezogen werden. Dass der Weltbegriff in der Tat auch in nicht-metaphorischen Kontexten recht vielfältig eingesetzt wird, wird anhand einer näheren Analyse des Wortes im ersten Teilkorpus deutlich. Längst nicht immer deutet der Kontext einer wörtlichen Verwendung von „Welt" auf die „große, weite Welt" hin. Oft wird auch ausgedrückt, dass etwas oder jemand eine bestimmte Qualität gegenüber anderen Dingen auf der Welt habe, häufig wird der großen, weiten Welt entsagt, indem die Einöde oder ein Kloster vorgezogen wird. Ebenfalls vielfältig ist der metaphorische Gebrauch des Wortes „Welt". Grob zusammengefasst wird der Begriff, abgesehen von der Referenz auf die gesamte geographische Welt, zur Bezeichnung von fünf unterschiedlichen metaphorisch-räumlichen Größen verwendet:

1. gesellschaftliche Welt oder sozialer Kreis,
2. die gesamte Menschheit,

[11] Eine tabellarische Übersicht der Daten zur Abbildung 8.6 befindet sich im GitHub-Repository mit den Forschungsdaten dieser Arbeit unter „Datenbasis der Analysen / Himmel und Welt / Welt / Welt-Annotationen Raabe-Kernkorpus" (Schumacher, 2021a).

3. Innenwelt,
4. mögliche Welt,
5. Personifizierung und Verdinglichung oder Vergleiche mit der Welt.[12]

In den Texten des 18. Jahrhunderts kommt außerdem die Redewendung „Welt haben“ oder „eine Person von Welt sein“ relativ häufig vor. Beim Vergleich der Kontexte des Weltbegriffs im 18. und 19. Jahrhundert fällt auf, dass hier grundsätzlich dieselbe Mehrdeutigkeit besteht. In beiden Jahrhunderten wird mit dem Begriff „Welt“ bei Weitem nicht nur auf die „große weite Welt“ Bezug genommen und selbst wenn, so geschieht dies nicht immer im Zusammenhang mit kolonialen oder Globalisierungserfahrungen. Häufig wird die Welt auch argumentativ (z. B. in Phrasen wie „das Beste auf der Welt“) eingebunden. Dass die Kontexte sich nicht grundsätzlich unterscheiden, bedeutet, dass diesbezüglich auch keine quantitativen Verzerrungen in Bezug auf den Weltbegriff als Störfaktor der oben ausgeführten Betrachtungen wirken.

Dass im Teilkorpus des 18. Jahrhunderts in sehr vielen Texten Raummetaphern vorkommen, die das Wort „Welt“ enthalten, illustriert eine Abfrage der Metaphern-datenbank *laRa* (vgl. Abbildung 8.7). Die Abfrage zeigt aber noch mehr: Es gibt in dem Teil des Trainingskorpus, der aus Ausschnitten von Erzähltexten des 18. Jahrhunderts zusammen gesetzt wurde, den größten Variantenreichtum von „Welt“-Metaphern. Sowohl die Verwendung des Wortes „Welt“ im metaphorischen Kontext als auch der Variantenreichtum nehmen über die Jahrhunderte stetig ab. Wie Abbildung 8.7 zeigt, kommt in 13 von 20 Texten des 18. Jahrhunderts im Trainingskorpus „Welt“ entweder als ein-Wort-Metapher oder als Teil einer metaphorischen Phrase oder eines komplexen Raumbildes in insgesamt 15 Varianten vor. Im Trainingskorpus des 19. Jahrhunderts kommt „Welt“ als Metapher noch in 12 Texten vor, die Varianten nehmen aber schon deutlich ab und sinken auf zehn. In dem Teil des Trainingskorpus, der aus Textausschnitten des 20. Jahrhunderts besteht, kommt „Welt“ als Metapher oder Teil einer Metapher nur noch in acht Texten vor und wieder nehmen die Varianten ab; es sind nur noch acht unterschiedliche „Welt“-Metaphern

[12] Beispiel 1.: „die gelehrte Welt“ (Jean Paul *Leben des Quintus Fixlein*); Beispiel 2.: „von aller Welt verlassen“ (Moritz *Andreas Hartknopf*, Schnabel *Wunderliche Fata einiger See-Fahrer*), „Indeß war die Welt weniger tolerant“ (Unger *Julchen Gruenthal*) Beispiel 3.: „mit ihrer inneren Welt gegen Kälte und Glut“ (Jean Paul *Leben des Quintus Fixlein*); Beispiel 4. „in dieser und in jener Welt“ (Miller *Siegwart*), „gleichsam in eine idealische Welt versetzt“ (Unger *Julchen Gruenthal*) „die glückliche Insel einer zweiten Welt“ (in Jean Paul *Hesperus oder die 45 Hundsposttage*), „noch vom Nebelstern der zweiten Welt“ (*Leben des Quintus Fixlein*). Beispiel 5: „Die Welt irrt“, „und eine ganze Welt lag auf seiner Seele“, „der moderne Leib ist eine Welt voll arbeitender Kräfte“ (alle drei in Jean Paul *Blumen-, Frucht- und Dornenstücke*).

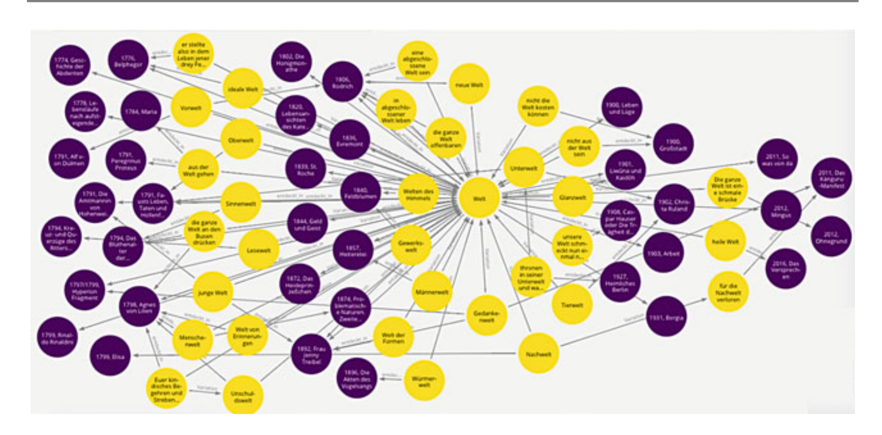


Abbildung 8.7 Raummetaphern mit Kernwort „Welt“ und deren Quellen

verzeichnet. Im Trainingskorpus des 21. Jahrhunderts konnte die metaphorische Verwendung des Wortes nur noch in fünf Texten ausgemacht werden. In diesem Teil des Trainingskorpus gibt es nur noch sechs Varianten (alle Variantenzahlen sind inklusive „Welt“ als ein-Wort-Metapher). Von diesen sechs Varianten ist eine besonders interessant, da sie eine Brücke schlägt zwischen einem Text des 21. Jahrhunderts und zwei Texten des 18. und 19. Jahrhunderts: Klings *Das Känguru-Manifest*, Fontanes *Frau Jenny Treibel* und Hölderlins *Hyperion Fragment* – in diesen drei Texten kommt (in den ersten 4.000 Tokens) die „Gedankenwelt“ vor. Auch die Metapher „Nachwelt“ verbindet das *Känguru Manifest* mit einem Text des 18. Jahrhunderts, nämlich *Elisa* von Wobeser. Weitere jahrhundertübergreifend genutzte Metaphern sind „Vorwelt“ in Nauberts *Alf von Dülmen* und de la Motte Fouqués *Rodrich* sowie „Unterwelt“ in Klingers *Faust Leben Taten und Höllenfahrt*, Dohms *Christa Ruland* und Hessels *heimliches Berlin* und „junge Welt“ in Nauberts *Amtmannin von Hohenweiler* und Fontanes *Frau Jenny Treibel*. Mit Ausnahme der ein-Wort-Metapher „Welt“ gibt es keine Metapher mit diesem Kernwort, die in allen vier Jahrhunderten vorkommt. Werden die Zahlenwerte des 18. Jahrhunderts in einen relationalen Wert umgewandelt, so ergibt sich daraus, dass in 65% der Texte im Trainingskorpus, die aus dem 18. Jahrhundert stammen, Welt-Metaphorik enthalten ist. Der Annahme folgend, dass ein ähnlich großer Anteil des ersten Teilkorpus den Ausdruck „Welt“ in einem metaphorischen Kontext enthält, müsste hier in 16 Texten „Welt“ eine metaphorische Bedeutung annehmen. Da die Stichprobe für eine ernsthafte Übertragung der Werte zu klein ist, wird diese Umrechnung allerdings lediglich herangezogen, um einen hypothetischen Vergleichswert zu erhalten,

der einen Hinweis darauf gibt, wie groß der Einfluss des metaphorischen Wortgebrauchs sein könnte. Betrachtet man nun die annotierten Instanzen des Wortes „Welt“ genauer und führt eine zusätzliche manuelle Annotation all jener Vorkommnisse durch, bei denen der direkte Kontext auf eine metaphorische Verwendung schließen lässt, so sind Rückschlüsse auf den tatsächlichen Anteil dieser Verwendungsform im Kernkorpus möglich. Im Teilkorpus des 18. Jahrhunderts wurden 650 der 1941 als „Ort“ annotierten Instanzen des Wortes „Welt“ als Metaphern einer der fünf oben genannten Kategorien interpretiert.[13] Das heißt, dass 33,5% der Wortvorkommnisse Metaphern sind oder innerhalb von metaphorischen Phrasen stehen. Der betrachtete Kontext erstreckt sich hier auf 10 Wörter vor und 10 Wörter nach dem annotierten Wort. Das Wort „Welt“ wird in diesem Teilkorpus in 24 von 25 Romanen (auch) metaphorisch genutzt. Im zweiten Teilkorpus wurden 488 der als Ort annotierten Vorkommnisse des Wortes „Welt“ als Raummetapher interpretiert. Das sind rund 35%, also ein ähnlich großer Anteil wie im ersten Teilkorpus. Wieder kommt der metaphorische Gebrauch des Wortes „Welt“ in 24 der 25 Romane in diesem Teilkorpus vor. Im dritten Teilkorpus wurden insgesamt 238 Annotationen von „Welt“ als Ort in 23 Romanen als Raummetaphern klassifiziert. Das macht 33,25% aller Annotationen von „Welt“ mit dem „Ort“-Tag aus. Im Teilkorpus des 21. Jahrhunderts wurden schließlich 280 der als „Ort“ annotierten Vorkommnisse von „Welt“ als Metapher bewertet. Das macht 26,17% der insgesamt 1.070 „Ort“-Annotation von „Welt“ aus. Der raummetaphorische Gebrauch des Wortes „Welt“ findet sich hier in 22 Romanen. Es ergibt sich also ein etwas anderes Bild als bei der Auswertung der Datenbank der Raum-Metaphern. Metaphern des Kernwortes „Welt“ kommen in allen vier Teilkorpora in einer eindeutigen Mehrzahl der Texte – nahezu in allen – vor. Der Anteil des metaphorischen Gebrauchs des Weltbegriffs an den gesamten Annotationen dieses Wortes bleibt vom 18. bis zum 20. Jahrhundert nahezu konstant und sinkt dann ab. Eine auffällige Gemeinsamkeit mit der Abfrage der Metaphern-Datenbank ist der Abfall des metaphorischen Gebrauchs von „Welt“ zum 21. Jahrhundert. Betrachtet man nun nicht den relativen Anteil sondern die absoluten Zahlen der Kontextanalyse, so wird erkennbar, dass der metaphorische Gebrauch tatsächlich einem ähnlichen diachronen Verlauf folgt wie die wörtliche Verwendung des Wortes. Daraus kann geschlossen werden, dass in der Tat der tatsächliche, das heißt in diesem Falle räumliche Gebrauch des Wortes „Welt“ Ausschlag gebend für die charakteristische Kurve des Weltbegriffs ist (vgl. Abbildung 8.8).

[13] Die Dokumentation dieser Kategorisierung befindet sich im GitHub-Repository der Forschungsdaten dieser Arbeit unter „Datenbasis der Analyse / Himmel und Welt / Welt / Kontexte gruppiert“ (Schumacher, 2021a).

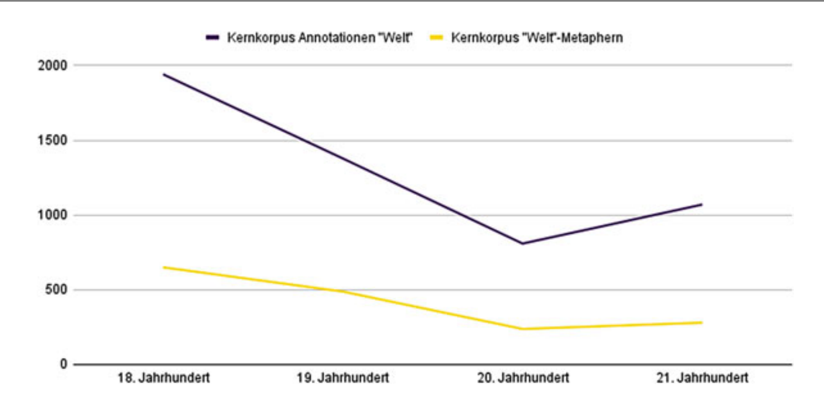


Abbildung 8.8 Quantitative Perspektiven auf den metaphorischen Gebrauch von „Welt“ im Kernkorpus

Wenn der Kontext auch nicht immer nahelegt, dass die Figuren sich in die Welt bewegen, sondern häufig z. B. auch, dass etwas sich von allen anderen seiner Art auf der Welt abhebt, so heißt das doch, dass die Welt in der Literatur des 18. Jahrhunderts häufiger argumentativ oder erzählerisch aufgerufen wird. Für die These, dass der Weltbegriff im 18. Jahrhundert eine besondere Bedeutung hat, die in den folgenden Jahrhunderten stark abfällt, spricht auch der Variantenreichtum des metaphorischen Gebrauchs dieses Begriffes im 18. Jahrhundert. Hier wird deutlich, dass der Weltbegriff nicht nur in seiner tatsächlichen Bedeutung besonders häufig genutzt wird, sondern, dass Autor*innen in diesem Jahrhundert auch besonders kreativ mit diesem Begriff umgehen. Sinne, Jugend, Erinnerungen, Unschuld – vielen nichträumlichen Größen wird hier zugesprochen, eine ganze Welt zu eröffnen. Die Daten stützen Krobbs These, dass das zweite Endeckungszeitalter den Weltbegriff auch ins Zentrum der Literatur rückt. Allerdings zeigt sich auch, dass dieser Trend ganz eindeutig im 18. Jahrhundert verankert ist. Einige der betrachteten Romane Wilhelm Raabes, fügen sich also nicht einfach nur in den Trend ihrer Zeit, sondern entsprechen im quantitativen Gebrauch des Weltbegriffes eher dem, was ein Jahrhundert zuvor üblich gewesen zu sein scheint.

Dadurch, dass hier auch zeitgenössische Texte im Vergleichskorpus betrachtet werden, zeigt sich außerdem ein interessanter aktueller Trend. Die Welt wird wieder häufiger aufgerufen und literarisch bearbeitet. Anders als im 18. Jahrhundert führt das aber nicht zu häufigerer metaphorischer Verwendung des Welt-Begriffes oder gar zu einer Steigerung des Variantenreichtums von Welt-Metaphern. Möglicherweise liegt das an einem thematisch nun nicht mehr an der Entdeckung orientierten Weltinteresse. Die Welt des 21. Jahrhunderts ist stark globalisiert. Eine andere Problematik

tritt hier aber erstmals im betrachteten Kernkorpus zutage: Der Begriff „Welt“ steht im Teilkorpus des 21. Jahrhunderts erstmalig explizit im Zusammenhang mit der Rettung derselben. Zwar findet sich eine direkte Kollokation von „Welt“ und „retten“ bzw. „Rettung“ (betrachtet wurde ein Wortumfeld von 3 Wörtern vor oder nach dem Begriff „Welt“) nur in zwei Romanen. Allerdings handelt es sich bei diesen Romanen genau um die beiden Ausreißer-Texte, die in Abbildung 8.6 sichtbar wurden, nämlich Steinhöfels *Mitte der Welt* und Bergs *Vielen Dank für das Leben*. Das Thema der Rettung der Welt könnte also mit der Häufung der Begriffsverwendung zusammenhängen, selbst wenn die Verwendung der Phrase „Welt retten“ in dieser expliziten Form nicht mehr als ein Einzelfall ist. Um eine solche Hypothese zu verifizieren, müsste allerdings ein ergänzendes Close Reading angeschlossen werden, dass nicht mehr Teil einer Fallstudie zum Weltbegriff bei Raabe sein kann.[14]

Fallstudie 2: Zur Kontextualisierung des Begriffs „Himmel“
Ein zweiter, kulturell stark aufgeladener Begriff, der eine ähnliche Entwicklungslinie aufzeigt wie „Welt“, ist der des Himmels. In dieser zweiten Fallstudie zu einer Einzelwortanalyse wird aber der umgekehrte Weg gewählt wie in der obenstehenden. War beim Weltbegriff die Frage leitend, inwiefern literaturwissenschaftliche Forschung auf quantitative Analysen aufbauen kann, so wird hier der Versuch unternommen, eine quantitative Betrachtung an analoge Forschungsansätze anzuschließen.

Anders als der Weltbegriff steht der Himmel nicht mit den Globalisierungsbewegungen, und damit einem Zeitphänomen, das direkt in den betrachteten Rahmen des Kernkorpus fällt, im Zusammenhang. Stattdessen stellen Oberdorfer und Waldow im Vorwort zu ihrem interdisziplinären Sammelband *Der Himmel als transkultureller ethischer Raum* fest, dass durch dessen Betrachtung „fundamentale Fragen nach dem Wert des menschlichen Lebens, der Stellung des Menschen im Kosmos und der Unterscheidbarkeit von Gut und Böse aufgeworfen“ (Oberdorfer und Waldow, 2016, 9) werden. Solche Fragen erscheinen zeitlos und keinen besonders großen Trend-Schwankungen unterworfen zu sein. Doch der Begriff des Himmels ist – ebenso wie der Weltbegriff – mehrdeutig. Der Himmel ist nicht nur Projektionsfläche menschlicher Selbstreflexion, sondern auch Forschungsgegenstand der Naturwissenschaften und Sitz des Göttlichen (vgl. Oberdorfer und Waldow, 2016, 10; Mieth, 2016, 105). Oberdorfer arbeitet am Beispiel von Heinrich Heines Literatur heraus, dass der Himmel durch religiöse Aufladung verknüpft ist mit dem

[14] Nebenbefunde wie dieser sind in Digital-Humanities-Forschung recht häufig und das Phänomen der Serendipity darum von besonderer Bedeutung. Genaueres dazu findet sich z. B. bei Schubert (2013).

Jenseits und damit einer späteren, post-Lebenszeit und mit den dann versprochenen ewigen Freuden (vgl. Oberdorfer und Waldow, 2016, 19; Giacobazzi, 2016, 250; Mieth, 2016, 105). Gleichzeitig sind der Himmel und damit auch die benannten Implikationen geographisch „oben" verortet (Oberdorfer und Waldow, 2016, 19). Wie bei Bachelard und Lotman stehen also auch hier das Oben und positive Wertung im Zusammenhang miteinander (vgl. Abschnitt 2.2.1). Im religiösen Kontext hat der Himmel zwei Konnotationen: Er ist der Aufenthaltsort Gottes und Verheißungsort des Menschen (Oberdorfer und Waldow, 2016, 21). Im metaphorischen Wortgebrauch bleibt vor allem die positive Aufladung enthalten und der Himmel steht dann nur noch für das Lichte, Helle, für das er in der religiösen Himmel-Hölle-Gegenüberstellung zum Inbegriff wurde (Oberdorfer und Waldow, 2016, 27). Am Beispiel von *Meister Eckhart* arbeitet Mieth heraus, dass der Himmel außerdem als Verinnerlichung des Göttlichen im Menschen selbst verortet werden kann (Mieth, 2016, 105). Bei der Analyse des Begriffs sind demnach sechs Bedeutungsebenen relevant:

1. Der physisch-geographische Himmel,
2. der Blick in den Himmel als Selbstreflexion,
3. der Himmel als Sitz Gottes, der Engel und anderer religiöser Größen,
4. der Himmel als Jenseitsort,
5. der interiorisierte Himmel,
6. der Himmel als Metapher für das Gute

Schaut man sich nur die quantitative Verteilung des Wortes „Himmel" im Kernkorpus an, so zeigt sich ein ganz ähnliches Bild wie beim Weltbegriff. Auch hier sinkt der Zahlenwert vom 18. zum 19. Jahrhundert um rund 50% von 1443 auf 676 ab. Vom 19. Jahrhundert bis zur Gegenwart gibt es dann eine stetige, aber nur leichte Abnahme auf 574 Annotationen im 20. und 498 Annotationen im 21. Jahrhundert. Um herauszufinden, ob der Gebrauch des Wortes „Himmel" insgesamt gleichmäßig abnimmt, oder ob einige der Kontextualisierungen von literarhistorischen Schwankungen betroffen sind, wurden die Annotationen des Wortes „Himmel" den sechs oben benannten, in der an literarischen Texten orientierten Forschung zum Himmel als ein transkulturelles, ethisches Raumphänomen herausgearbeiteten, Kontext-Kategorien zugeordnet.[15] Es ergibt sich das in Abbildung 8.9 dargestellte Bild.

[15] Die Daten zu dieser Fallstudie sind im GitHub-Repository zur vorliegenden Arbeit im Ordner „Datenbasis der Analyse / Welt und Himmel / Himmel" abgelegt und können dort heruntergeladen werden (Schumacher, 2021a).

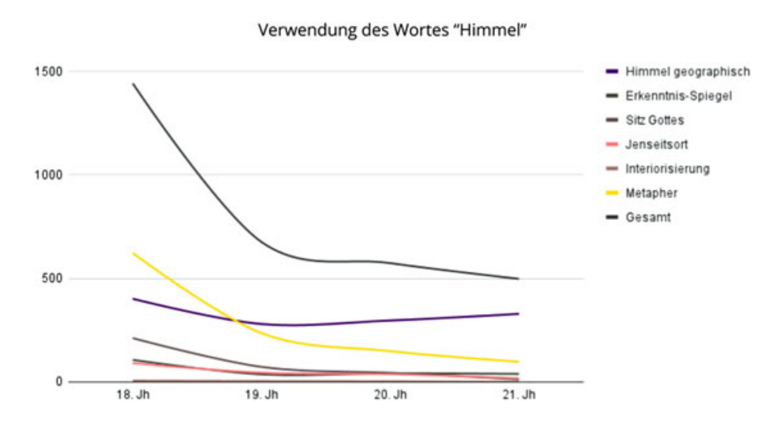


Abbildung 8.9 Nach Kontext zugeordnete Bedeutungskategorien des Himmelsbegriffs im Zeitverlauf

Nur eine der Kontext-Kategorien des Wortes „Himmel“ weist eine leichte Schwankung, bzw. keine deutliche Abnahme im diachronen Verlauf auf: die des Himmels als physisch-geographischer Größe. Himmel als Metapher, als Sitz Gottes und Jenseitsort, sowie der Himmel als Erkenntnis-Spiegel und Interiorisierung des Himmels finden im Zeitverlauf immer weniger Erwähnung. Dieser Fund passt zu der Annahme, dass in der Moderne der Himmel nicht mehr einstimmig als Sitz Gottes angesehen wird, sondern dass Gott zunehmend in Irdischem gesehen wird (Giacobazzi, 2016, 250). Interessant ist auch der Sonderfall der Interiorisierung des Himmels. Auch wenn eine solche Kontextualisierung insgesamt sehr selten ist, so kommt sie im ersten Teilkorpus doch immerhin sieben Mal vor, im zweiten immerhin noch fünf Mal und im 20. und 21. Jahrhundert schließlich nur noch vier und ein Mal. Führt man beide Beobachtungen zusammen, so lässt sich die These aufstellen, dass sich Religiosität vom 18. bis zum 21. Jahrhundert in literarischen Texten immer weniger widerspiegelt. Nicht nur wurde der Himmel als Jenseitsort und Sitz Gottes im 18. Jahrhundert weit häufiger erwähnt als in den folgenden, auch wurden religiöse Ideen, wie die bei Meister Eckhart beobachtete Interiorisierung des Himmels, hier hin und wieder aufgegriffen. Auch bei der metaphorischen Verwendung des Himmelsbegriff spielt Religiosität eine große Rolle. Wenn der Himmel z. B. personifiziert wird, so steht er oft für eine göttliche Macht und selbst die Gleichsetzung von Himmel mit einem positiven Wert kann darauf zurück geführt werden, dass der Himmel als Sitz Gottes und paradiesischer Jenseitsort diesen positiven Wert begründen. All diese Kontexte gehen in der Literatur des 18. bis 21. Jahrhunderts stark zurück. Dagegen bleibt das physisch-geographische Himmelsverständnis für

den literarischen Raum in etwa konstant wichtig, was vor allem Schilderungen des Wetters, Beschreibungen von Größenverhältnissen oder die Verortung von fliegenden Tieren betrifft. Der Blick in den Himmel als Erkenntnis-Spiegel wird zwar im 18. Jahrhundert auch auffallend häufig erwähnt, die Zahlen bleiben dann aber vom 19.–21. Jahrhundert relativ konstant. Hier zeigt sich die zeitlose Faszination des Menschen für die unendlich weite und unerreichbare Größe, die, wie Oberdorfer und Waldow es beschrieben haben, dazu beiträgt, den Menschen zu fundamentalen Lebensfragen anzuregen (Oberdorfer und Waldow, 2016, 9). Auch in dieser Fallstudie wurde zum Abgleich die Datenbank *laRa* nach Raummetaphern des Kernwortes „Himmel“ abgefragt. Wie Abbildung 8.10 zeigt, kommen in den Anfangspassagen von 14 von 20 Romanen (70%) des 18. und 13 von 20 Romanen (65%) des 19. Jahrhunderts Metaphern vor, deren Kernwort „Himmel“ ist. In dem Teil des Trainingskorpus, das sich aus Romanen des 20. Jahrhunderts zusammensetzt, sind es zwei (10%) und in dem aus dem 21. Jahrhundert drei (15%).

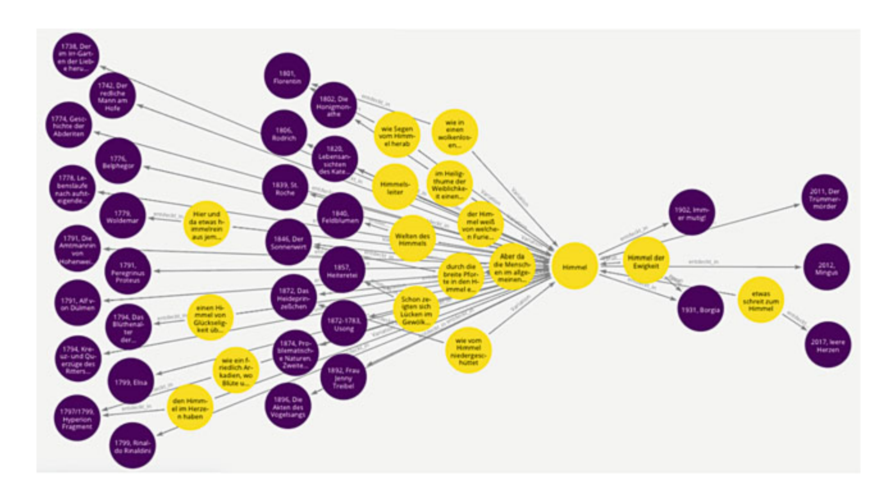


Abbildung 8.10 Raummetaphern mit Kernwort „Himmel“ und deren Quellen

Es zeigt sich also ebenfalls ein deutlicher Abwärtstrend. Vor allem angesichts der relativ kleinen Stichprobe im Trainingskorpus, die zur Basis der Datenbank der Raummetaphern wurde, überrascht es, dass die Daten der Klassifizierung von Annotationen des Ausdrucks „Himmel“ als Metapher im Kernkorpus relativ ähnlich sind wie die der Datenbankabfrage. Im ersten Teilkorpus stehen 623 der als „Ort“ annotierten Vorkommnisse des Wortes in einem metaphorischen Kontext, also rund

43% der insgesamt 1442 Annotationen. Diese Wortvorkommnisse befinden sich in 23 der 25 Romane im ersten Teilkorpus (92%). Im zweiten Teilkorpus wurde „Himmel“ 236 Mal als Raummetapher interpretiert, also 35% Prozent aller „Himmel“-Annotationen und diese Annotationen erstrecken sich auf 24 der 25 Romane (96%). Im dritten Teilkorpus sind es dann 149 (26%) in 22 Romanen (88%) und im vierten 98 (19,7%) in 20 Romanen (80%). Datenbankabfrage und Analyse des Kernkorpus zeigen also ein ähnliches Bild. Es gibt einen deutlichen Abwärtstrend vom 18. bis zum 21. Jahrhundert. Durch die genaue Betrachtung der einzelnen Annotationen im Kernkorpus, die sich jeweils auf ganze Romane beziehen (während in *laRa* das Trainingskorpus aufbereitet wurde, das aus den Romanen jeweils nur 4.000-Token-Abschnitte enthält), wird sichtbar, dass im 21. Jahrhundert zwar in vielen Romanen „Himmel“ in einem metaphorischen Kontext steht, dass aber die Anzahl der Verwendung innerhalb der einzelnen Romane stark abnimmt. Sowohl in Bovenschen *Sarahs Gesetz* als auch in Herrndorf *Bilder einer großen Liebe* oder Uebel *Horror Vacui* steht jeweils nur ein Mal das Wort „Himmel“ in einem metaphorischen Kontext.[16] Außerdem zeigt *laRa* noch zwei weitere Besonderheiten der Himmels-Metapher, die vor allem im Vergleich mit der metaphorischen Verwendung des Weltbegriffs auffallen: Der Variantenreichtum ist im 19. Jahrhundert am höchsten und die Varianten sind nicht nur korpus- sondern nahezu ausnahmslos textspezifisch. Abgesehen von „Himmel“ als ein-Wort-Metapher wird keine Metapher jahrhundertübergreifend genutzt.

[16] Im Hinblick auf Best Practices einer reflektierten digitalen Textanalyse (vgl. dazu die Beiträge des von Reiter, Pichler und Kuhn (2020) herausgegebenen Sammelbandes *Reflektierte algorithmische Textanalyse*, insbes. (Pichler und Reiter, 2020, 43–60; Kuhn, 2020, 61–106 und Gerstorfer, 2020, 107–124) sei hier darauf hingewiesen, dass sich im Laufe der vorliegenden Studie herausgestellt hat, dass es von großer Bedeutung ist, die Stärken und Schwächen des jeweiligen Forschungssettings von (Teil-)Analysen und die daraus resultierende Perspektive auf die betrachteten Phänomene genau im Blick zu behalten. Die Betrachtung der relativen Werte der Verwendungen von Raummetaphern im Kernkorpus ist sehr gut geeignet, um in die Tiefe zu gehen und z. B. wie hier die unterschiedlichen Verwendungszusammenhänge zu betrachten. Dabei sollte aber nicht vergessen werden, dass die Romane, die hier in ihrer Gesamtlänge betrachtet werden, in der Wortanzahl stark variieren (je länger eine betrachtete Passage ist, desto höher ist auch die Wahrscheinlichkeit, dass darin eine der häufig genutzten Raummetaphern „Welt“ oder „Himmel“ vorkommt). Dieser Faktor wird in einer Korpusgestaltung wie derjenigen, die zum Fundament für *laRa* wurde, minimiert. Die Anzahl der Tokens pro Roman und der Romane pro Jahrhundert werden konstant gehalten und nur in seltenen Fällen sind Autor*innen mit mehr als einem Text vertreten. Auf diese Weise wird möglichst hohe Vergleichbarkeit der Teilabschnitte des Trainingskorpus gewährleistet. In den Abfragen zeigen sich so trotz einer relativ kleinen Datenbasis von kurzen Ausschnitten aus 80 Texten allgemeine diachrone Trends der Verwendung von Raummetaphern in literarischen Texten.

Insgesamt kann nach diesen zwei Fallstudien-basierten Tests von *laRa* festgehalten werden, dass die relationale Datenbank der Raummetaphern schon jetzt, also mit relativ kleinem Datenumfang, ein gutes Werkzeug ist, mit dem eine digitale Textanalyse sinnvoll ergänzt werden kann. Die Stärke des auf Vergleichbarkeit der einzelnen Teilabschnitte des Korpus optimierten Systems liegt darin, allgemeine Trends der Verwendung von Raummetaphern aufzuzeigen und den Variantenreichtum und die damit verbundene Vernetzung sichtbar zu machen. Auf diese Weise können der ungefähre quantitative Einsatz von Metaphern eines bestimmten Kernwortes im literarhistorischen Kontext eingeschätzt und Rückschlüsse auf die kreative Verarbeitung einzelner Begriffe gezogen werden.

Mit „Welt" und „Himmel" wurden hier zwei für die literarische Darstellung von Raum quantitativ bedeutsame Wörter betrachtet, die vergleichsweise häufig metaphorisch genutzt werden. In den beiden Fallstudien wurde gezeigt, wie literaturwissenschaftliche Analysen quantitative Ansätze nutzen können, um einen betrachteten Fokus-Gegenstand in größere literarhistorische Kontexte einzubetten oder wie quantitative Analysen an die nicht digital unterstützte literaturwissenschaftliche Forschung anknüpfen können, um z. B. weitere Kategoriensysteme bei der Sichtung von Wortvorkommnissen einzusetzen. Die Verbindung von digitalen und analogen Ansätzen kann sich auf diese Weise sinnvoll ergänzen. Auch der Abgleich mit der Metaphern-Datenbank *laRa* hat sich als gute Methode erwiesen, um die Abfrage-Ergebnisse der Raum-Annotationen kontextualisieren zu können. Obwohl die Datenbank auf dem Trainingskorpus, also einer relativ kleinen Datenbasis beruht, stimmen die Abfrage-Ergebnisse tendenziell mit den Auswertungen der Klassifizierungen der Raum-Annotationen im Kernkorpus überein oder ergänzen diese. Die Datenbank kann also bei Fallstudien wie den hier beispielhaft durchgeführten bereits entscheidende Hinweise auf den metaphorischen Gebrauch einzelner Wörter in literarischen Texten des 18.–21. Jahrhunderts geben. Ein genauerer Abgleich vermittelt aber einen umfassenderen Einblick in die tatsächlich zum Kernkorpus gehörenden Erzähltexte. Was ein digitaler Ansatz nicht unterstützt, sind Untersuchungen, die konzeptionell den Himmel nicht als Begriff im Text fokussieren, sondern ein metaphorisches Himmelsverständnis zur Basis machen und als Indikatoren dafür implizite Hinweise auf Glück, Liebe und Fantasiewelten ausmachen, wie z. B. Giacobazzi in seinem Aufsatz *Die Dialektik von Himmel und Erde zwischen Klassik, Romantik und Realismus* (2016: 249–260).

8.1.5 Der literarische Raum ist ein System der Öffnung und Schließung

Die Betrachtung der Raumthemen hat gezeigt, dass eines von ihnen von überragender Bedeutung ist: das Thema der Grenzen bzw. der Öffnung und Schließung von Räumen. Dieses Thema hatte bereits Lotman (1973) in seinem Modell literarischen Raumes als bedeutsam ausgemacht. Es ist später aber auch in weniger semiotischen und stärker funktional ausgerichteten narratologischen Ansätzen zum literarischen Raum aufgegriffen worden. Bei Ryan et al. wird beschrieben, auf welche Weise Öffnung und Schließung von Räumen die Handlung vorantreiben oder verlangsamen kann (vgl. Ryan et al., 2016, 22 und Abschnitt 2.2 dieser Arbeit). Auch Bal (1985) beschreibt, inwiefern literarische Räume der Öffnung und Schließung mit Themen wie Mobilität und auch mit Figurencharakterisierungen verknüpft sein können. Dabei stellt sie z. B. fest, dass männliche Figuren häufiger beweglich sind, während weibliche eher in statischen Räumen bleiben Bal and van Boheemen-Saaf, 1985, 137. Nun ist der Raum-Classifier zwar darauf trainiert, Wörter zu erkennen, die zu Raumthemen gehören, er weist aber keine Unterkategorien zu. Das hat den großen Vorteil, dass das Machine-Learning-Tool auf diese Weise stärker auf variablen Kontext reagiert und z. B. auch Wörter annotiert, die – aus dem Kontext geschlossen – zu Raumthemen gehören, für die es aber (noch) keine Einzelkategorie gibt. An dieser Stelle zeigt sich deutlich, warum für diese Studie ein Machine-Learning-Ansatz gegenüber listenbasierten Ansätzen bevorzugt wurde. Der Classifier kann durch seine Kontextsensitivität Wörter annotieren, die – aus dem Kontext geschlossen – zu einem Raumthema gehören könnten, die aber nicht von vornherein als Indikatoren für Raumthemen berücksichtigt wurden. Auf diese Weise können auch im diachronen Verlauf sich neu herausbildende Raumthemen erfasst werden. Anders herum könnten aber aus den Annotationen des Raum-Classifiers umfangreiche Listen mit Wörtern erstellt werden, die zu einzelnen Raumthemen gehören, die dann mit Hilfe dieser Listen weiter betrachtet werden können, z. B. in anderen Korpora. Der Nachteil der hier gewählten Kategorisierung ohne weitere Unterkategorien ist, dass die Daten des Raum-Classifiers erst noch einmal weiter klassifiziert werden müssen, bevor sie aussagekräftig werden. Hier ist die automatische Raum-Annotation also ein reines Preprocessing, das für die weitere literaturwissenschaftliche Analyse weiterbearbeitet werden muss. Darum wurden die Wörter, die vom Raum-Classifier als Raumthemen zugehörig annotiert wurden, in einem interpretativen Prozess weiter klassifiziert. Dabei wurden die Tabellen mit allen mit dem Tag „Raumthema" versehenen Wörtern der vier Teilkorpora einzeln ausgewertet. Aus den in diesen Tabellen quantitativ zusammengefassten Types wurden zuerst diejenigen gruppiert, die klar einem der Raumthemen zugeordnet werden konnten, die in der

literaturwissenschaftlichen Forschung bisher hervorgehoben wurden (vgl. Abschnitt 2.2.2). Neben Reise, Abenteuer und der Thematik der Öffnung und Schließung gehören dazu auch die heterotopen Räume des Theaters bzw. der Bühne, des Friedhofs und allem, was mit Bestattungsriten verbunden ist sowie des Spiegels als heterotop-utopisches Objekt. Für Wörter, die nicht einer dieser Kategorien zugeordnet werden konnten, wurden eigene Kategorien gefunden, wie z. B. Freiheit, Krieg, Einsamkeit, Flucht/Zuflucht, Ordnung, Begegnung, Dasein, und Natur/Landleben/Jagd. Insgesamt wurden auf diese Weise 27 Raumthemen identifiziert. Drei dieser Raumthemen kamen im Teilkorpus des 19. Jahrhunderts hinzu (Suche, Weite, Heimat), drei weitere (Einkauf, Tarnung und Fernsehen) im 21. Jahrhundert.[17]

Ganz deutlich zeigt die Übersicht in Tabelle 8.5 die quantitativ überragende Bedeutung des Themas „Öffnung und Schließung“. Zu diesem Thema gehören Wörter wie „Grenze“, „Tür“, „Graben“, „Fenster“, „Treppe“ u. Ä., die sehr häufig deutlich machen, ob ein Innenraum im Fokus steht oder ob eine Figur sich gerade außen aufhält oder ob die Grenzen (z. B. durch die bereits in ihrer Bedeutung hervorgehobenen Blickachsen) überbrückt werden. Der deutliche quantitative Abstand zwischen den beiden am häufigsten aufgerufenen Themen und den folgenden kann zum Teil auch dadurch erklärt werden, dass die semantischen Felder dieser Themen sehr groß sind und sehr viele Wörter (Types) umfassen. Doch auch die nicht in Themen zusammengefassten Rohdaten der Abfrage des Tags „Raumthema“ untermauern den Schluss, dass für literarische Raumthematik die Öffnung / Schließung nach Lotman (1973) besonders wichtig ist. Die am häufigsten mit dem Tag „Raumthema“ belegten Wörter sind nämlich tatsächlich „Fenster“ (2.556) und „Tür“ (2.199).

Die Visualisierungen der Raumthemen[18] in den einzelnen Teilkorpora (vgl. Abbildung 8.11) zeigen deutlich, dass das Thema der Öffnung und Schließung (größter heller Kasten) zwar durchgehend von großer Bedeutung ist, dass sich aber eine diachrone Entwicklung abzeichnet. Während im 18. Jahrhundert das Thema Reise (größter dunkler Kasten) noch häufiger aufgerufen wurde (Öffnung/Schließung: 3778, Reise: 5211), so nimmt es in den folgenden Jahrhunderten immer weniger der literarischen Raumdarstellung ein. Im 19. und 20. Jahrhundert gibt es jeweils etwa doppelt so viele Annotationen, die zum Raumthema der Öffnung/Schließung

[17] Welche Wörter zu welchem Raumthema klassifiziert wurden, zeigen die Tabellen im Ordner „Datenbasis der Analyse / Raumthemen“, die im GitHub-Repository zu dieser Studie abgelegt wurden (Schumacher, 2021a).

[18] Das vergleichsweise kleine Format eines Buches führt bei dieser Abbildung zu einer unlesbaren Schriftgröße. Im GitHub-Repository (Schumacher, 2021a) zu dieser Studie im Ordner „Abbildungen von ‚Orte und Räume im Roman‘“ befinden sich die vier Thementafeln daher einzeln als SVG-Dateien, die beliebig vergrößert werden können.

Tabelle 8.5 Auflistung der Raumthemen und zugeordnete Wortanzahl im Kernkorpus

Raumthema	Anzahl
Öffnung / Schließung	21.429
Reise	13.377
Freiheit	2.611
Begegnung	2.181
Krieg	1.822
Flucht / Zuflucht	724
Natur / Landleben / Jagd	650
Spiegel	546
Einsamkeit	458
Abenteuer	440
Erscheinung	420
Ordnung	389
Leere	186
Aufführung	179
Bestattung	179
Kirche / Räume der Spiritualität	110
Suche	106
Räume der Krankheit	81
Naturkatastrophen / Unfälle	72
Dasein	59
Fest	37
Weite	34
Heimat	13
Einkauf	13
Tarnung	9
Fernsehen	5

gehören, wie zum Thema der Reise. Im 21. Jahrhundert wurden fast drei Mal so viele Wörter dem Grenzthema zugeordnet wie der Reisethematik.

Interessant ist aber auch, dass im gesamten Kernkorpus Wörter, die für Öffnung stehen wie z. B. „öffnete" (986 Annotationen), „offen" (755 Annotationen), „öffnen" (330 Annotationen), „geöffnet" (319 Annotationen) oder „offenen" (275 Annotationen) häufiger sind als solche, die eindeutig Schließung ausdrücken wie

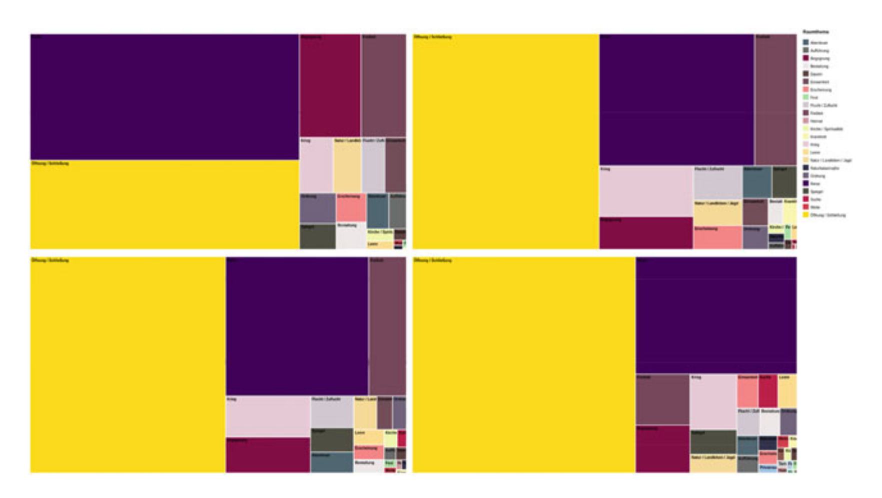


Abbildung 8.11 Raumthemen-Tafeln vom 18. (links oben) bis zum 21. Jh. (rechts unten)

z. B. „geschlossen“ (325 Annotationen) oder „schloß“ (169 Annotationen) bzw. „schloss“ (130 Annotationen). Die Darstellungen literarischen Raumes sind also zwar durchzogen von Grenzen, diese stellen sich aber häufiger als permeabel denn als unüberwindbar dar. Raum wird häufiger eröffnet und somit Mobilität ermöglicht; Grenzen sind meist überwindbar.

Die absolute Anzahl der Wörter, die als Raumthemen aufrufend annotiert wurden, schwankt stark. Waren es im 18. Jahrhundert 10.820, so steigt die Anzahl im 19. Jahrhundert auf 14.402. Dann folgt im 20. Jahrhundert ein Tiefstand von 9.047. Im 21. Jahrhundert wurden dann wieder 10.804 Wörter als zu Raumthemen gehörig annotiert. Die Anzahl der Types folgt hier einer ähnlichen Wellenbewegung, wenn die Amplitude auch deutlich schwächer ausfällt. Die Vielfalt der Wörter, die als zu Raumthemen gehörig klassifiziert wurden, bleibt mit 885 Types im 18. Jahrhundert, 1.121 im 19. Jahrhundert, 986 im 20. Jahrhundert und 989 im 21. Jahrhundert immer auf einem ähnlichen Niveau. Dies legt nahe, dass Raumthemen über einen relativ klar definierten Wortschatz aufgerufen werden, der sich zwar über die Jahrhunderte verändert, aber nur in relativ geringem Ausmaß.

Die Betrachtungsweise der Raumthemen, die durch die automatische Annotation des Raum-Classifiers eingeleitet wird, entspricht ziemlich genau dem Distant-Reading-Ansatz nach Moretti (vgl. Moretti, 2013). Damit geht einher, dass hier in geringerem Maße Anschlussfähigkeit an literaturwissenschaftliche Einzeltext- oder Einzelphänomen-Analyse gegeben ist. Auch wenn die Annotationen weiter unter-

klassifiziert und dann in Gruppen genauer betrachtet werden, ist dies kein Schritt zurück in die literarischen Texte, so wie es z. B. die Kontextanalyse in Abschnitt 8.1.4 war. Eigentlich wird die Distanz zum Einzeltext durch die Gruppierung sogar noch größer, da von den Einzelwörtern, die im Prinzip noch auf die Kontexte zurückführbar wären, weiter abstrahiert wird und am Ende Wortgruppen im Zentrum der Betrachtung stehen. Auch wenn die Anbindung an Close-Reading-Analysen in diesem Falle schwierig ist, so ist stattdessen die Verknüpfung mit konzeptuellen Überlegungen zum literarischen Raum wie denen von Lotman (1973), Bal (1985) oder Ryan et al. (2016) sehr gut möglich. An diese kann nämlich mit allgemeinen Tendenzen angeschlossen werden. Wenn Ryan et al. anhand eines Beispiels illustrieren, inwiefern der literarische Raum die Handlung eines Erzähltextes bestimmt, wenn Figuren, wie z. B. Odysseus beim Kampf gegen die Freier seiner Frau, räumliche Grenzen gesetzt oder Wege eröffnet werden (Ryan et al., 2016, 42), so kann nach der quantitativen Analyse der Raumthemen festgestellt werden, dass die Öffnungen literarischen Raums insgesamt überwiegen. Auf diese Weise kann das gezeichnete Bild eines dynamischen Raums, der am Ende aber eine grundsätzliche Offenheit und Überwindbarkeit von Grenzen suggeriert, abgerundet werden.

8.1.6 Die zunehmende Privatheit literarischen Raumes

Immer wieder wurde in diesem Kapitel die Frage nach der Anschlussfähigkeit der hier angewendeten Methode an traditionellere Ansätze der Literaturwissenschaften gestellt. Es wurde bereits gezeigt, dass diese z. B. durch Einzelwortanalysen von Begriffen, die für bedeutungshaltige Konzepte stehen, durch Betrachtungen einzelner Raumkategorien wie die der Raumthemen und durch die Anknüpfung sichtbar gewordener Strukturen an konzeptionelle Überlegungen, hergestellt werden kann. Nun soll noch ein weiterer Fall von Anschlussfähigkeit gezeigt werden, nämlich der einer Raumkategorien-übergreifenden Analyse des in den Literaturwissenschaften eher diffus betrachteten Konzeptes der Privatheit. Bevor aber dafür argumentiert wird, dass die Darstellung von Privatheit im diachronen Verlauf zunimmt, muss klar festgelegt werden, wie Privatheit verstanden werden kann.

Im 2018 herausgegebenen Sammelband *Privates Erzählen. Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts* von Burk, Klepikova und Piegsa werden durchaus sehr unterschiedliche Aspekte von Privatheit herausgearbeitet. Grundsätzlich werde diese durch Grenzen konstituiert, die aber überschreitbar sind (Burk, 2018a, 11). Die Thematik der Privatheit kann also als Unterthema der Lotman'schen Grenzthematik klassifiziert werden. Klar ist auch, dass Privatheit mehrere Dimensionen aufweist und auf unterschiedliche Weise betrachtet werden

kann. Besonders anschlussfähig für digitale Herangehensweisen scheint ein Konzept von Rössler, in dem zwischen informationeller, dezisionaler und lokaler Privatheit (Rössler, 2001) unterschieden wird und das sowohl auf Manns *Buddenbrooks* als Roman des 20. (Burk, 2018b, 91–128) als auch auf Überwachungsromane des 21. Jahrhunderts (Huber, 2018, 195–218) gewinnbringend angewendet wurde. Von den drei Aspekten wird der Fokus dieser Analyse entsprechend auf die lokale Privatheit gelegt. Dass die anderen Aspekte aber durchaus in die digitale diachrone Analyse einbezogen werden können, wird am Rande berücksichtigt. Sehr häufig wird die lokale Privatheit mit Innenräumen verknüpft, die allerdings in ihrer Privatheit – je nach kulturhistorischem Zusammenhang – stark variieren können. Im 18. Jahrhundert kann der Innenraum grundsätzlich von Aspekten des Öffentlichen eingenommen werden (Goeth, 2018, 54). Im Idealfall wird er zu dieser Zeit aber zum privaten Rückzugsraum (Lehnert, 2014, 17). Im 19. Jahrhundert gibt es im Philistertum eine geradezu bühnenartige Inszenierung häuslicher Räume, eine echte Privatheit kann hier nicht erreicht werden (Czezior, 2018, 85–86) und auch im 20. Jahrhundert kann eine zu stark auf Repräsentativität ausgerichtete Häuslichkeit nur scheitern (Burk, 2018b, 106ff). Während eine Verortung im lokal privaten Bereich der Wohnung zwar Privatheit markieren kann, so ist doch das wahrhaft Private das diskret Verborgene, wie Dröscher-Teille (2018, 297–314) es am Beispiel von Bachmanns *Malina* zeigt. In Überwachungsromanen des 21. Jahrhunderts werden private Räume so stark ausgeleuchtet, dass Alternativen z. B. in der Natur gefunden werden müssen (Huber, 2018, 204–207). Scheitert Privatheit in Räumen, die diese eigentlich gewähren sollten, so fliehen die beschriebenen literarischen Figuren häufig – und das zieht sich durch viele Beiträge des oben genannten Sammelbandes – in eine private Innenwelt, die im Extremfall eine Fantasiewelt ist (vgl. Knode, 2018, 39 ff.; Goeth, 2018, 62–63; Czezior, 2018, 82ff; Huber, 2018, 209–210). Auch wenn in literaturwissenschaftlichen, meist auf Einzeltexte oder sehr kleine Korpora beschränkten, Fallstudien viele weitere Aspekte herausgearbeitet werden, so lässt sich doch übergreifend und zusammenfassend festhalten, dass räumliche Privatheit im Idealfall im gesamten häuslichen Wohnumfeld verortet werden kann. Kann sie es nicht, so dringt die Öffentlichkeit zunächst in Räume ein, die als weniger privat empfunden werden, und erst dann in geschütztere Bereiche. Dadurch werden die zuvor privaten Räume dieser Qualität beraubt und in repräsentative Räume verwandelt. Besondere Schutzzonen sind Räume, die mit zwischenmenschlicher Intimität verbunden werden, insbesondere das (weibliche) Schlafzimmer und hier vor allem das Bett (Passavant, 2018, 136). Noch privater sind die Innenwelten der Figuren, die in Abschnitt 8.1.2 bereits thematisiert wurden und darum hier nicht noch einmal ausführlich aufgegriffen werden.

Nun hat sich ein Bewusstsein über die Privatheit von (Innen-) Räumen kulturgeschichtlich erst im 18. Jahrhundert herausgebildet (Lehnert, 2014, 19). Das Korpus dieser Analyse ist darum ein idealer Gegenstand für eine etwas breiter angelegte diachrone Betrachtung dieses Themas. Die Analyse der Types, die zum literarisch bedeutsamsten Raumthema der Öffnung und Schließung (wie im vorangegangenen Kapitel beschrieben) beitragen, hat bereits gezeigt, dass Innen- bzw. Wohnräume und deren Abgrenzung einen wichtigen Teil zu diesem Raumthema beitragen. Privatheit ist ein literarisches Motiv, das sich nicht klar in eine der sechs Raumkategorien einordnen lässt. Es handelt sich zudem um eine skalare Kategorie, da etwas mehr oder weniger privat sein kann. Einige Zimmer, die hier als Ort klassifiziert werden, können z. B. privater sein als andere. Privatheit kann zum Raumthema der Öffnung und Schließung gehören, es kann aber auch ein eigenes Thema werden, wie die Tatsache zeigt, dass im Teilkorpus des 21. Jahrhunderts insgesamt 21 Wörter als einem Raumthema zugehörig annotiert wurden, die eindeutig die Privatheit dieser Räume implizieren (Beispiele sind „Schlafzimmertür" oder „Badezimmertür") und die darum als Privaträume klassifiziert wurden (vgl. Abbildung 8.12). Diese Zahl ist sehr klein und zeigt, dass spezifische Themen sich sehr langsam herauskristallisieren. Erstmals wurde ein einzelnes Token dem Thema „Privatraum" im Teilkorpus des 19. Jahrhunderts zugeordnet. Im dritten Teilkorpus waren es dann 2 Wörter, die als zu diesem Unterthema gehörig klassifiziert wurden. Diese Zahlen berücksichtigt, kann ein Sprung auf 21 Wörter schon als Indikator dafür betrachtet werden, dass sich hier ein neues Unterthema herausbildet.

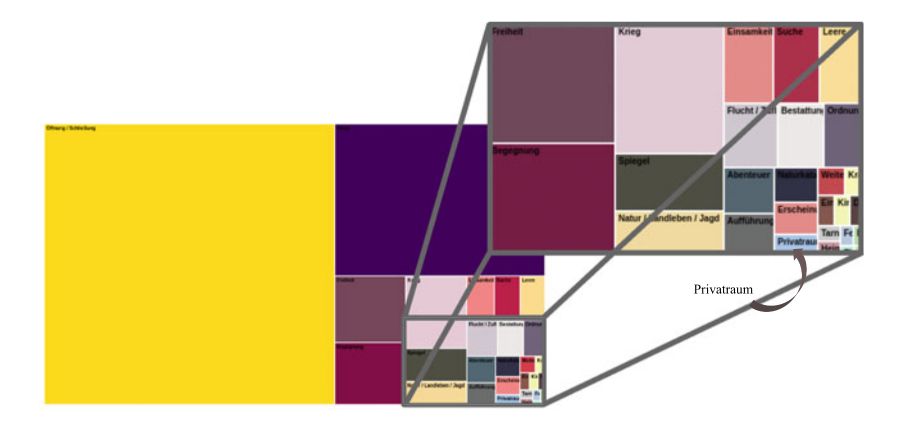


Abbildung 8.12 Thementafel des 21. Jahrhunderts mit Zoom auf „Privatraum"

Objekte, die hier meist als Raumhinweise interpretiert werden, können ebenfalls zu Symbolen der Privatheit werden. Am explizitesten weisen aber Raumbeschreibungen auf Öffentlichkeit und Privatheit hin, denn hier werden die Adjektive „privat" und „öffentlich" einsortiert, die ich nun als erstes betrachte.

Im Teilkorpus des 18. Jahrhunderts steht „öffentlich" noch relativ weit oben auf der Rangliste der am häufigsten als Raumbeschreibung annotierten Wörter.[19] Der gegenteilige Ausdruck „privat" spielt hier aber eine noch untergeordnete Rolle. Gegen 167 Annotationen von „öffentlich", 113 von „öffentlichen" und 43 von „öffentliche" stehen insgesamt 16 Vorkommnisse von Wörtern, die mit „privat" beginnen im gesamten Teilkorpus. Eine Abfrage der Wörter, die mit „privat" oder „Privat" beginnen, zeigt 96 Ergebnisse. Keines dieser Vorkommnisse wurde vom Raum-Classifier als „Raumbeschreibung" klassifiziert. Auch im Teilkorpus des 19. Jahrhunderts wurde „privat" nicht als Raumbeschreibung kategorisiert. Eine mögliche Ursache ist, dass die Privatheit, die sich kulturgeschichtlich gerade erst herausbildet (s.o.), zu diesem Zeitpunkt noch selten explizit ihren Weg in die Literatur findet. Eine andere Möglichkeit ist, dass das Konzept der Privatheit in den früheren der betrachteten Jahrhunderte noch nicht unbedingt mit Raum verbunden wird. Ergänzende Wortabfragen in CATMA und anschließende Kontextanalysen führen diesbezüglich zu genaueren Einsichten.

Im Teilkorpus des 19. Jahrhunderts ist die attributive Verwendung „privat" noch seltener als im zuvor betrachteten. Eine CATMA-Wortabfrage zeigt, dass nur zwölf Wörter mit „privat" beginnen. Automatisch annotiert wurde davon wieder keines. Um festzustellen, ob es sich hier um einen Fehler des Raum-Classifiers handelt, wurden die Kontexte des Lemmas „privat" analysiert. Bei einer Abfrage der Wörter, die mit „privat" oder „Privat" beginnen (insgesamt 81), hat sich gezeigt, dass dieser Wortstamm in der Tat nicht immer raumbeschreibend eingesetzt wird. Stattdessen wird es häufig synonym zu „persönlich" eingesetzt, wie z. B. in der Phrase „privater Standpunkt" (in Fontane *Unwiederbringlich*) oder „private Wohltätigkeit" (in Goedsche *Sebastopol*). Im Teilkorpus des 18. Jahrhunderts ist das Verb „privatisieren" (bei Jean Paul in *Blumen-, Frucht- und Dornenstücke* oder in *Hesperus oder die 45 Hundsposttage*) und die Bezeichnung „Privatperson" (in Schnabel *Fata einiger See-Fahrer*) noch vergleichsweise häufig, beide Ausdrücke bezeichnen eher eine Opposition zum Beruflichen denn zum Öffentlichen. Die Abfrage des Wortbeginns „Privat" zeigt ein ähnliches Ergebnis. Auch hier werden selten lokale Aspekte von Privatheit referenziert, stattdessen wird ebenfalls häufig das

[19] Die Abfragen der Annotationen der Kategorie „Raumbeschreibung" pro Teilkorpus befinden sich im GitHub-Repository der Forschungsdaten zu dieser Arbeit unter „Datenbasis der Analyse / Raumbeschreibung", die der Abfragen zur Privatheit und Öffentlichkeit im Unterordner „Privatheit" (Schumacher, 2021a).

nicht-Berufliche bezeichnet oder informationelle Privatheit (z. B. als „Privatangelegenheiten“ oder „Privatgespräch“) angesprochen. Dass diese Types nicht vom Raum-Classifier erkannt wurden, ist also kein Fehler des Tools, sondern zeigt tatsächlich, dass Privatheit in der Literatur des 18. und 19. Jahrhunderts noch nicht in erster Linie ein Raumphänomen war. Nach Rössler (2001) überwiegt hier also die informationelle die lokale Privatheit. Die Abfrage der Lemmata „Öffentlich“ und „öffentlich“ zeigt, dass sie im Teilkorpus des 18. Jahrhunderts 405 Mal vorkommen. Davon wurden 323 vom Raum-Classifier erkannt. Im Teilkorpus des 19. Jahrhunderts ergibt dieselbe Suche 172 Ergebnisse, davon wurden 128 vom Raum-Classifier gefunden und klassifiziert. Im Gegensatz zur Privatheit wird das Öffentliche auch im 18. und 19. Jahrhundert häufig schon mit Raum verknüpft. Sowohl im ersten als auch im zweiten Teilkorpus findet das Öffentliche häufiger explizit Erwähnung als das Private. Hier zeigt sich außerdem ein Vorteil der auf maschinellem Lernen basierenden Erkennung und Klassifizierung gegenüber Listen-basierten Ansätzen. Der Raum-Classifier erkennt Begriffe, die mit „privat“ oder „öffentlich“ beginnen tatsächlich meist nur dann als Indikatoren für Raumdarstellungen, wenn der Kontext auf eine solche Verwendung hinweist. Die hier durchgeführten listenbasierten Abfragen eignen sich, wie gezeigt, aber hervorragend, um einen Vergleich der klassifizierten Indikatoren für Raumthemen mit den nicht automatisch annotierten Wörtern durchzuführen. Auf diese Weise lasen sich literarhistorische Entwicklungen von Begriffsverwendungen erkennen und in die Analyse einbeziehen.

Auch im Teilkorpus des 20. Jahrhunderts wurde „privat“ nicht als Raumbeschreibung klassifiziert. Das Lemma „öffentlich“ findet sich nur noch 22 Mal unter den automatisch generierten Annotationen von Raumbeschreibungen. Eine analog zu den für die ersten beiden Teilkorpora durchgeführte Abfrage der Wortanfänge „Privat“ und „privat“ führt 61 Ergebnisse auf. Die Abfrage der Wortanfänge „Öffentlich“ und „öffentlich“ ergibt nur 53 Treffer. Erstmalig geht hier also die explizite Kennzeichnung von etwas als öffentlich gegenüber der Klassifizierung von etwas als privat zurück. Für das Teilkorpus des 21. Jahrhunderts zeigt sich ein ähnliches Bild. Unter den Annotationen der Raumbeschreibungen findet sich keine, die explizit Privatheit bezeichnet, dafür aber 54 Annotationen des Lemmas „öffentlich“. Eine direkte Abfrage der Wortanfänge „Privat“ und „privat“ ergibt aber 106 Tokens, von „Ö/öffentlich“ 115 Tokens. Jenseits der lokalen Privatheit werden Privatheit und Öffentlichkeit also wieder in etwa gleich häufig aufgerufen. Dabei fällt auch auf, dass die explizite Referenzierung von Öffentlichkeit über die Jahrhunderte seltener wird. Aber wie verhält es sich mit privaten Räumen, die auch vom Raum-Classifier erkannt werden können? Die These, dass die Darstellung von Privatheit auch in ihrer räumlichen Dimension im 21. Jahrhundert einen Höhepunkt erreicht, wird nicht nur dadurch gestützt, dass sie zum ersten Mal als Raumthema ausfindig gemacht werden

kann, sondern auch noch durch ein ganz anderes Phänomen jenseits der Raumbeschreibungen.

Erstmalig steht im Teilkorpus des 21. Jahrhunderts ein Objekt unter den ersten zehn am häufigsten annotierten Wörtern der Kategorie „Raumhinweis". „Bett" hat es hier auf Platz zehn geschafft.[20] Und damit ist die Privatheit, die im 18. Jahrhundert so langsam anfing in die Literatur einzudringen, nun endgültig darin angekommen. Denn die Kontexte zeigen, dass Figuren sich nicht nur ins Bett zurückziehen, um einen Ort für zwischenmenschliche Intimität zu haben, sondern sehr häufig auch, wenn Erschöpfung sich breit macht oder ein ruhiger Ort zum Nachdenken gesucht wird. Das Bett ist also Ort der Intimität, der Erholung und des Überlegens und meistens auch des Rückzugs und der Einsam-, höchstens Zweisamkeit. Das Bett wird in diesem Teilkorpus aber nicht nur mit insgesamt 801 Annotationen besonders häufig zum Raumhinweis, die Annotationen erstrecken sich auch auf das gesamte Korpus, also alle 25 Romane. Hinzu kommt, dass auch der Körper als potentieller Indikator dezisionaler Privatheit im Sinne einer letzten Rückzugsmöglickeit in die Innenwelt (Huber, 2018, 207–208) in diesem Teilkorpus häufiger als in den zuvor betrachteten als Raumhinweis annotiert wurde.[21] Wenn das Bett hier für objektgebundene Privatheit steht, so wird der Körper zum subjektgebundenen Pendant. Auch in den Kontexten des Wortes „Körper" geht es um private Eigenheiten und Geschehnisse. Geschildert werden Einzelkörper oder Zweisamkeiten. Der Körper als physische Manifestation einer Figur wird sowohl intern als auch extern gelesen, d.h. es werden die Empfindungen von Figuren über ihren Körper, über Schmerzen und Verfall z. B. dargestellt, aber auch wie andere Figuren Körper von außen wahrnehmen, als schön oder hässlich, gesund oder ruiniert. Der Körper, der in der literaturwissenschaftlichen Forschung ebenso wie in der Philosophie geradezu als Nullpunkt der Wahrnehmung von Raum ausgemacht wurde (vgl. Abschnitt 2.2 und 2.3), wird in der Literatur des 21. Jahrhunderts zum wichtigen Einschreibungsraum (im Sinne von Deleuze, 2006 vgl. Abschnitt 2.3.3) von Innen- und Außenwelten. Der letzte Rückzugsraum in die Privatheit (s.o.) scheint ebenso vom Verfall bedroht wie die privaten Außenräume. Denn Innen- und Außenwelt stehen in einer Wechselbeziehung. Wie diese beschaffen ist, das zeigt der Körper als Grenze zwischen Innen und Außen. So heißt es z. B. in Zaimoglus *Isabel* „Arme Christine, tolle Christine, Körper kaputt, Seele futsch." und es bleibt offen, ob nun die Zerrüttung der Seele

[20] Dieses Teilergebnis kann anhand der Abfragen der Annotationen der Kategorie „Raumhinweis" nachvollzogen werden, die sich im GitHub-Repository der Forschungsdaten unter „Datenbasis der Analyse / Raumhinweis" befinden (Schumacher, 2021a).

[21] Die Kontexte der Annotationen von „Bett" und „Körper" können im Forschungsdaten-Repository auf GitHub unter „Datenbasis der Analyse / Raumhinweis / Bett und Körper" eingesehen werden (Schumacher, 2021a).

der des Körpers voran ging oder umgekehrt. In einer Passage aus Steinhöfels *Die Mitte der Welt* ist hingegen klar, dass Krankheit, und damit ein Faktor im Inneren, den Körper prägt. Dort heißt es: „Der Körper ist so ausgemergelt, dass man dessen kantige Umrisse unter der bis an die Brust reichenden Bettdecke kaum ausmachen kann." Auch in Trojanows *Weltensammler* gibt es eine Passage, in der eindeutig die innere Beschaffenheit, dieses Mal allerdings keine physiologische, sondern eine psychische, einer Figur sich in deren Körper einschreibt: „Als sie sein Gehöft erreichen, liegt er [es geht um einen König; Anm. M.K.S] im Schatten des königlichen Baums, ein Körper, aufgedunsen über jedes Maß hinaus, ein Anführer, der keinerlei Bewegung schätzt." In einem anderen Roman, *Über Nacht* von Sabine Gruber, wird die Verbindung von Körper und Raum sogar explizit thematisiert: „Aber im selben Moment musste sie daran denken, dass sie wohl nicht aufhören würde, den eigenen Körper als Nullpunkt eines Koordinatensystems wahrzunehmen, mit dessen Hilfe sie alles um sich herum einteilte". Während also das Bett ein Objekt im privaten räumlichen Umfeld von Figuren ist, ist der Körper Ausdruck subjektiver Innen- und Außenwelten einer Figur bzw. von deren Zusammenspiel. Schilderungen und Erwähnungen von beiden tragen zur Darstellung sehr privater Räume bei, wie sie in den vorigen Jahrhunderten noch nicht zu finden waren.

Auch wenn nicht zwingend eine Abnahme der Wörter „öffentlich" und „Öffentlich" im Zusammenhang mit der Zunahme von Darstellungen privater Räume bis hin zur Schilderung von Intimität als extremer Form der Privatheit stehen muss, so fällt doch auf, dass es eine sinkende Tendenz gibt. Die Anzahl der Vorkommnisse mit Wörtern dieses Wortstamms nimmt vom Teilkorpus des 18. zum Teilkorpus des 19. Jahrhunderts um mehr als die Hälfte ab (von 405 auf 172). Dann sinkt der Wert im Teilkorpus des 20. Jahrhunderts auf einen Tiefstand von 53 und steigt im 21. Jahrhundert wieder auf 106 an. Die quantitative Bedeutung der früheren Jahrhunderte wird also nicht mehr erreicht. Es ist anzunehmen, dass die zunehmende Privatheit nicht nur für sich als Thema immer wichtiger wird, sondern dass diese die Öffentlichkeit als häufig aufgerufenem literarischen Raum (nahezu) ablöst.

Es könnten noch zahlreiche weitere vom Raum-Classifier annotierte Wörter, die implizit oder explizit auf Privatheit hindeuten, in einer tiefergehenden Analyse betrachtet werden. Inwiefern eine auf dem Preprocessing mit Hilfe der automatischen Annotation aufbauende digitale Untersuchung bisherige Ansätze der Literaturwissenschaften ergänzen kann, wird in diesen beispielhaften Betrachtungen aber bereits deutlich. Anhand eines skalaren Verständnisses von Privatheit, das Wohnräume generell, intime Wohnräume insbesondere und Innenwelten als ultimativ privat umfasst, kann hier diachron nachgewiesen werden, dass eine Verlagerung von den weiter außen liegenden Wohnräumen zum Innersten der Privatheit der figürlichen Innenwelt stattgefunden hat. Darüber hinaus konnte gezeigt werden, dass diese

Innenräume ebenfalls an Privatheit verlieren, wenn sie sich über die Grenzen des Körpers nach außen zeigen. Was mit Hilfe dieser Methode nicht oder nur schwer erfasst werden kann, ist die kulturelle Umdeutung von Räumen als privat oder intim, wie sie z. B. Huber für die Natur im Überwachungsroman feststellt (Huber, 2018, 204–207) und wie sie auch schon in der Literatur der Romantik stattgefunden hat (vgl. Huber, 2018, 203; Doering, 2013, 352–353). Um den diachronen Verlauf im Blick behalten zu können, wird hauptsächlich mit festen Indikatoren auf Wortebene operiert und doch lässt die Kontextsensitivität des Machine-Learning-Verfahrens es zu, in begrenztem Umfang den Wandel eines eher diffus verstandenen Konzeptes wie das der Privatheit zumindest in Teilaspekten in einem vergleichsweise großen Korpus zu untersuchen und damit Gedanken aus analogen Ansätzen der Literaturwissenschaft weiterzuführen.

8.2 Perspektive: jahrhundertspezifische Teilkorpora

Anhand der allgemeinen Tendenzen im gesamten Kernkorpus wurde in Abschnitt 8.1 betrachtet, inwiefern die mit dem Raum-Classifier generierten Daten Thesen aus nicht-digitalen literaturwissenschaftlichen Studien ergänzen, stützen oder manchmal auch widerlegen können. Es ist aber auch möglich, die digitale Methode eher explorativ einzusetzen. Dabei werden besondere Häufungen und Relationen zu anderen Teilkorpora in den Fokus genommen. Ein solcher Zugang wird in diesem Kapitel nun anhand der korpusspezifischen Phänomene beispielhaft gezeigt, um auch mit dieser Herangehensweise das Potential der Methodik auszuloten.

Das diachrone Projektdesign dieser Studie bringt die Schwierigkeit mit sich, im ersten Teilkorpus Besonderheiten und spezifische Phänomene auszumachen. Stattdessen legt die Betrachtung der Daten des 18. Jahrhunderts eher den Grundstein, zeigt die Vergleichswerte für die folgenden Analysen. So hat sich bei der Auswertung der Datenbasis aus dem 18. Jahrhundert schon deutlich die Relationalität literarischer Raumdarstellung gezeigt. Auch die Gegenüberstellung von äußeren und inneren Räumen ist im ersten Teilkorpus bereits von großer Bedeutung. Die ebenfalls schon in Abschnitt 8.1.4 geschilderte diachrone Bedeutungsabnahme des Weltbegriffes nimmt ihren Anfang im 18. Jahrhundert. Im Umkehrschluss ist also die relative Häufigkeit des Wortes „Welt“ ein Spezifikum des ersten Teilkorpus. Aufgrund dieser perspektivbedingten Schwierigkeit, korpusspezifische Phänomene im ersten Teilkorpus zu identifizieren und um Dopplungen zu vermeiden, liegt der Schwerpunkt der Betrachtungen dieses Abschnitts eher auf den späteren Jahrhunderten. Eine kleinere, bisher unerwähnt gebliebene Beobachtung konnte aber auch für das 18. Jahrhundert ausgemacht werden. So ist dies das einzige Teilkorpus,

in dem das Thema der Reise häufiger aufgerufen wird als das der Öffnung und Schließung. Die Themen Freiheit und Krieg sind im 18. Jahrhundert zwar ebenfalls besonders häufig. Das gilt aber auch und ganz besonders für das Teilkorpus des 19. Jahrhunderts. Zwar bleibt das Kriegsthema auch im Teilkorpus des 20. Jahrhunderts häufig, steht aber wieder weit zurück hinter dem Thema der Freiheit. Außerdem ist interessant, dass im Teilkorpus des 20. Jahrhunderts erst- und einmalig die nationalstaatliche Raumbeschreibung „deutsch“ unter den fünf meistannotierten Wörtern der Kategorie „Raumbeschreibung“ steht. Im 21. Jahrhundert nehmen schließlich die Besonderheiten zu. Das liegt zum einen wieder daran, dass die früheren Jahrhunderte eher als Ausgangspunkt dienen, zum anderen ist aber ein nicht zu unterschätzender Faktor die mangelnde Kanonisierung dieser Texte.[22] Auch wird hier ein vergleichsweise kurzer Zeitraum literarischen Schaffens abgedeckt, nämlich statt 100 nur rund 20 Jahre, da das 21. Jahrhundert zum Zeitpunkt der Entstehung dieser Studie nicht viel mehr als diese Zeitspanne umfasst. Die Texte liegen also zeitlich näher beieinander und sind gleichzeitig weniger homogen, was die Ansprüche an kanonisierte Texte angeht. Die Besonderheiten, die sich hier zeigen, sind, dass mehr heterotope Räume beschrieben werden, dass die Thematik der Leere sich zunehmend als eigenes Raumthema herauskristallisiert, dass Ausdrücke der „Fremde“ eine semantische Umdeutung erfahren und dass körperliche Beschreibungen bzw. optische Klassifizierungen von Figuren als „dick“ oder „dünn“ viel häufiger sind als in den anderen drei Teilkorpora.

[22] Während es sich bei den Texten des 18.–20. Jahrhunderts zumeist um solche handelt, die bereits häufig Teil literaturwissenschaftlicher Studien waren, in Literaturgeschichten genannt werden oder Teil der Schulausbildung sind und aufgrund dieser Tatsachen zum literarischen Kanon gehören, bilden sie auch einen Teil literarischer Veröffentlichungen ab, der – aus unterschiedlichen Gründen – als besonders wertvoll gilt (zu literaturwissenschaftlichen Wertungstheorien vgl. Rippl und Winko, 2013, 25–49). Dass immer wieder auf bestimmte Werke zurückgegriffen wird, kann ästhetische (Freise et al., 2013) aber auch inhaltliche Gründe, wie den Aufbau eines kulturellen Gedächtnisses (Assmann, 2013), und weitere haben. Die Texte des 21. Jahrhunderts haben aufgrund ihrer Aktualität solche Prozesse noch nicht durchlaufen. Nur einzelne dieser Romane, wie z. B. Stanišić *Vor dem Fest* (vgl. z. B. Huber, 2017; Sandberg, 2018; Römer, 2018 und Freudenstein, 2016, 203–204), sind bereits Teil von Forschung, Literaturgeschichten und schulischer Allgemeinbildung. Die stilistische und thematische Ausrichtung dieses Teilkorpus ist also beliebiger, da hier aus einer größeren Fülle von Veröffentlichungen ausgewählt wurde. Wie genau der Kanonbias der ersten drei Teilkorpora beschaffen ist, kann nicht im Rahmen dieser Studie untersucht werden (einige, wenn auch nur wenige, Aspekte zur Kanonbias im, dem Kernkorpus zu Grunde liegenden, TextGrid-Repository zeigen Dziudzia und Hall, 2020 auf). Dass dieser aber beim vierten Teilkorpus fehlt, muss im Hinblick auf die größere Vielfalt korpusspezifischer Phänomene angemerkt werden.

8.2.1 Spezifika des ersten Teilkorpus (18. Jahrhundert)

Im Teilkorpus des 18. Jahrhunderts ist das Raumthema der Reise mit 5.211 annotierten Wörtern deutlich präsenter als das der Öffnung und Schließung (3.678 Annotationen). Damit wird dieses Thema nicht nur in diesem Teilkorpus deutlich häufiger aufgerufen als das Grenzthema, das Lotman (1973) als besonders bedeutsam für literarische Raumdarstellung beschrieben hat, sondern auch sehr viel häufiger als in den folgenden Jahrhunderten. Mit 3.541 Annotationen im zweiten, 2.184 im dritten und 2.441 im vierten Teilkorpus zeigt sich insgesamt ein Abwärtstrend. Wie ist nun aber das semantische Feld dieses literarischen Themas im 18. Jahrhundert beschaffen? Auf diese Frage möchte ich kurz eingehen.

Die am explizitesten zum Raumthema der Reise beitragenden Wörter – Reise, reisen, Abreise u. ä. – sind auch die häufigsten. Gefolgt werden sie von Ausdrücken, die die Art des Transportmittels verraten. Dabei werden „Schiffe“ und „Wagen“ am häufigsten erwähnt, die „Kutsche“ und der „Zug“ folgen erheblich später. Häufig sind außerdem Wörter, die den Zustand des Auf-Reisen-Seins mehr oder weniger explizit beschreiben, wie z. B. „unterwegs“ oder auch „Aufenthalt“. Die häufigsten Wörter dieses Themas sind also insgesamt relativ allgemein gehalten und beschreiben die Reise nicht näher. Unter den nur ein Mal als Raumthema annotierten Wörtern, die der Reise zugeordnet wurden, finden sich zwar einige, die die Reise etwas plastischer machen, wie z. B. „Kur“ oder „Badereise“, „Reisecoffre“ oder „Equipage“, doch auch hier findet man viele allgemeine Ausdrücke wie z. B. ungewöhnlichere Schreibweisen eines häufigen Ausdrucks (Beispiele hierfür sind „Rükreise“ und „Rükkunft“). Das semantische Feld des Reisethemas, welches vom Raum-Classifier ausgemacht werden kann, ist also beherrscht von recht unspezifischen und insgesamt eher wenig herausragenden Ausdrücken. In den späteren Jahrhunderten ändert sich das ein wenig. Hier rücken Fahrzeuge, im 19. und 20. Jahrhundert vor allem die Begriffe „Wagen“ und „Schiff“, im 21. Jahrhundert dann auch das „Auto“, deutlich in den Vordergrund. Es ist später also nicht mehr nur wichtig, dass man unterwegs ist, sondern auch wie. Auch wenn die Unterschiede in der Ausgestaltung des Reisethemas nicht sehr groß sind, so zeichnet sich doch eine Entwicklung ab. Im 18. Jahrhundert stand die Reise allgemein im Mittelpunkt, Fahrzeuge wurden ebenfalls häufig benannt, allerdings war insgesamt bedeutender, dass man reiste und nicht die spezielle Ausgestaltung einer Reise. Ab dem 19. Jahrhundert wurde die Ausgestaltung der Reise dann wichtiger, Fahrzeuge wurden häufiger erwähnt. Dass sich das Reisen auch durch technologischen Fortschritt verändert, zeigen die Erwähnungen unterschiedlicher Fahrzeuge von der Kutsche bis zum Auto. Interessanterweise ist für die literarische Reisethematik in allen vier Jahrhunderten das Schiff von besonderer Bedeutung, ein Fahrzeug, das zumindest im

zeitgenössischen Alltag seltener zur Fortbewegung genutzt wird als andere. Möglich wäre, dass das Schiff in literarischen Texten bis zur Gegenwart immer wieder benannt wird, da es ein Fortbewegungsmittel ist, das bereits in der Antike mit dem Topos der Reise verknüpft war (wie z. B. in Homers *Odyssee*). Hier zeigt sich deutlich, dass die in dieser Studie angewendete Methodik ihre Stärken im Identifizieren von meist in hoher Anzahl gebrauchten Wörtern hat, die die Basis eines spezifischen Themas bilden können, dieses aber nicht voll ausdifferenzieren oder gar erzählerisch ausschmücken. An dieser Stelle wäre ein ergänzendes Close Reading sinnvoll, wenn der Fokus auf einem bestimmten Thema liegt und nicht – wie hier – darauf, eine beispielhafte Exploration der Daten zu zeigen.

8.2.2 Spezifika des zweiten Teilkorpus (19. Jahrhundert)

Dass räumliche Phänomene sich im ersten und zweiten Teilkorpus meist sehr stark ähneln, wurde bereits erwähnt. Das Teilkorpus des 19. Jahrhunderts weist allerdings ein Spezifikum auf, die es deutlich von dem des 18. Jahrhunderts und auch von den folgenden abgrenzt. Im 19. Jahrhundert wird nämlich die Thematik des Krieges vergleichsweise häufig und auf besonders positiv-beschreibende Weise aufgerufen. „Krieg“ ist hier nach „Öffnung und Schließung“, „Reise“ und „Freiheit“ das Raumthema, das die vierthäufigste Wortanzahl in diesem Teilkorpus umfasst. Dass mit 958 Annotationen das Thema der Freiheit sehr ähnlich häufig ist, ist ebenfalls interessant, da die Themen inhaltlich nah beieinander liegen („Gefangenschaft“ wird z. B. zum Thema „Freiheit“ klassifiziert, kann aber auch im Kontext des Krieges stehen). Mit insgesamt 818 Wörtern, die das Kriegsthema im Teilkorpus des 19. Jahrhunderts stützen, hat das Thema ungefähr die dreifache Wortanzahl wie im Teilkorpus des 18. Jahrhunderts (hier war „Krieg“ mit 287 Wörtern auf Platz fünf). Das Raumthema der Freiheit ist quantitativ (721 Wörter im 18. und 958 im 19. Jahrhundert) in den ersten beiden Teilkorpora dagegen viel näher beieinander. Die Bedeutungszunahme des Themas Krieg zeigt, dass hier regionale und nationale Auseinandersetzungen ins Zentrum rücken. Zu der Tatsache, dass das Kriegsthema hier so häufig aufgerufen wird, passt auch, dass Flucht/Zuflucht ein Thema ist, dass mit 289 Annotationen auf Platz sechs der Raumthemen steht. Deutlich an Bedeutung verliert das Thema des Krieges hingegen im 20. Jahrhundert. Hier umfasst es nur noch 380 Annotationen. Das überrascht zunächst, da das 20. Jahrhundert das Jahrhundert zweier Weltkriege war. Nun ist die Auswahl dieses Teilkorpus per Zufallsgenerator erfolgt und es sind tatsächlich – obwohl zumindest die Zeit des ersten Weltkriegs von den Veröffentlichungsdaten abgedeckt wird – keine zum Genre der Kriegsromane gehörigen Erzähltexte dabei. Die Zeit des zweiten Weltkriegs wird nicht abgedeckt,

da die Werke nur weniger Autoren, die zu dieser Zeit und kurz danach veröffentlicht haben, bisher gemeinfrei und darum in digitalen Archiven zugänglich gemacht worden sind. Es ist also möglich, dass hier ein ganz anderer Eindruck entstünde, wenn dezidiert Kriegsliteratur mit aufgenommen worden wäre.[23] Doch auch in den Werken, die sich nicht explizit dem Thema Krieg widmen, zeigt sich das Thema insgesamt als bedeutend für die Raumdarstellung. Dass das Thema der Freiheit eine ähnliche Entwicklungslinie zeigt (mit 721 Annotationen im 18. Jahrhundert sehr präsent, dann im 19. Jahrhundert mit 958 Annotationen am häufigsten und dann mit abnehmender Tendenz zum 20. mit 572 Annotationen und 21. Jahrhundert mit 360 Annotationen), deutet darauf hin, dass die beiden Themen eng zusammenhängen. Denn gerade in Kriegszeiten (ebenso wie in Vor- oder Nachkriegszeiten) kann die Sehnsucht nach Freiheit präsent sein und gerade in Kriegszeiten ist auch Gefangenschaft häufig. Im Teilkorpus des 21. Jahrhunderts nehmen die Wörter, die zu den Themen Freiheit und Krieg beitragen, weiter ab. Trotzdem umfassen die Themen Freiheit (358) und Krieg (344) weiterhin die dritt- und viertmeisten Wörter der Raumthemen. Hier wirkt einerseits das Weltkriegsjahrhundert noch nach, auf der anderen Seite zeigen sich aber auch Tendenzen zur Globalisierung. Neben den Weltkriegen finden hier z. B. auch Krimkrieg und Koreakrieg Erwähnung. Die Bedeutung des Themas Flucht/Zuflucht nimmt aber im Teilkorpus des 21. Jahrhunderts ab.

Die beschriebenen Tendenzen und semantischen Verschiebungen zeigen sich auch, wenn die einzelnen Wörter, die in einem Teilkorpus der Kriegsthematik zugeordnet wurden, auf Grundlage ihrer jeweiligen Anzahlen in Form einfacher Wortwolken visualisiert und nebeneinandergestellt werden, wie in Abbildung 8.13.



Abbildung 8.13 Wortwolken der Kriegsthematik in den vier Teilkorpora

[23] Allerdings muss hier auch darauf hingewiesen werden, dass auch in den anderen drei Teilkorpora die Auswahl der Texte nach einem möglichst neutralen Prinzip stattgefunden hat und auch hier kein Genre gegenüber anderen bevorzugt wurde. Dadurch bleibt die Vergleichbarkeit der Teilkorpora untereinander besser gewährleistet auch wenn dafür in Kauf genommen wird, dass literarhistorische Genre-Häufungen unberücksichtigt bleiben.

Hier ist schnell ersichtlich, dass im zweiten Teilkorpus die meisten Wörter (Types) zur Thematik des Krieges beitragen. Aber auch qualitative Aspekte werden sichtbar. Im ersten Teilkorpus spielt der Türkenkrieg eine Rolle. Es geht um „Angriff", „Einfall", „Belagerung", „Untergang" und „Zerstörung". Aber auch von „Revolution" ist die Rede. Neben vergleichsweise neutralen, beschreibenden Ausdrücken fallen hier negativ geprägte Wörter wie „Untergang" besonders ins Auge. Das Vokabular des 19. Jahrhunderts scheint einerseits militärisch-beschreibender, andererseits positiver. Hier ist der „Kriegsrath" als häufiges Wort erkennbar, ebenso wie „Kriegsherren", „Kriegsschiff" oder „Armee". „Schlacht", „Angriff" und „Kampf" sind ebenfalls besonders häufig erwähnt. Als positiv besetztes Wort fällt hier der „Sieg" besonders auf. Im dritten Teilkorpus nimmt die Wortanzahl wieder sichtbar ab. Hier ist der Ausdruck „Angriff" ausgesprochen häufig. Daneben stehen aber auch deutlich negativ konnotierte Wörter wie „Untergang", „Rückzug", „Greuel" oder „Ausbruch" unter den häufigsten Ausdrücken. Auch hier wird der „Angriff" wieder vergleichsweise häufig erwähnt. Das Vokabular nimmt im 21. Jahrhundert wieder deutlich ab. Wieder werden „Angriff" und „Schlacht" häufig genannt. Daneben fallen negativ konnotierte Wörter besonders auf, wie z. B. „Untergang", „Einfall", „Kriegsgefangenschaft" oder „Kriegsgebiet". Wenn auch längst nicht so häufig wie im 19. Jahrhundert (15 Mal), so wird doch auch hier „Sieg" als positiv konnotierter Begriff noch erwähnt (vier Mal). Außerdem tritt hier das Wort „Weltkrieg" in den Vordergrund. Von einem eher negativ geprägten semantischen Feld entwickelt sich das Kriegsthema also im 19. Jahrhundert nicht nur zu einem in der Literatur häufig vorkommenden Raumthema, sondern auch zu einem vergleichsweise neutral beschriebenen und zum Teil sogar positiv konnotierten. Im 20. Jahrhundert gibt es sowohl positiv als auch negativ konnotierte Begriffe die häufig sind. Im 21. Jahrhundert schließlich ist das Kriegsthema dann eher eines, das negativ beschrieben wird. Hier zeichnet sich also nicht nur eine quantitative Besonderheit des 19. Jahrhunderts ab, man kann auch sehen, dass in diesem literarhistorischen Abschnitt das Kriegsthema am positivsten dargestellt wird.

8.2.3 Spezifika des dritten Teilkorpus (20. Jahrhundert)

Das dritte Teilkorpus weist bei den Raumbeschreibungen einen interessanten Peak auf. Hier ist erst- und einmalig das Adjektiv „deutsch" (das hier im Anschluss an die automatische Annotation manuell lemmatisiert wurde) auf Platz fünf der häufigsten

Raumbeschreibungen.[24] Im ersten Teilkorpus war es mit 235 Annotationen auf Platz elf, im zweiten mit 251 auf Platz zwölf und im vierten Teilkorpus mit nur noch 116 Annotationen auf Platz 19. Auch das semantische Umfeld der nationalstaatlichen Zuschreibungen zeigt interessante Veränderungen im Laufe der Jahrhunderte.

Im 18. Jahrhundert werden vergleichsweise selten Nationalitäten genannt. Am häufigsten sind hier mit 235 Annotationen die Wörter, die sich auf das Lemma „deutsch“ rückführen lassen (betrachtet werden hier immer die ersten zehn Plätze nationalstaatlicher Raumbeschreibungen). Danach kommt mit 74 Annotationen „englisch“. Es folgen absteigend „italienisch“ (60), „griechisch“ (49), „französisch“ (35), „preußisch“ (16), „polnisch“ (12), „spanisch“ (7), „holländisch“ (6) und „russisch“ (6). Es ist ein sehr westlich-nachbarschaftliches Umfeld, das hier am häufigsten aufgerufen wird. Im 19. Jahrhundert ist überraschenderweise „englisch“ mit 320 Annotationen auf dem ersten Platz. Erst dann folgt mit 251 Annotationen „deutsch“. Dahinter stehen auf Platz drei bis zehn „griechisch“ (213), „französisch“ (190), „russisch“ (163), „polnisch“ (64), „italienisch“ (60), „orientalisch“ (39), „preußisch“ (38) und „türkisch“ (33). Nicht nur weitet sich das geographische Umfeld hier in Richtung Osten und Südosten deutlich aus, es fällt auch auf, dass die absoluten Zahlenwerte der zehn am häufigsten aufgerufenen nationalstaatlichen Zuordnungen viel höher sind als im ersten Teilkorpus. Dieser Faktor relativiert sich im dritten Teilkorpus wieder. Dafür steht hier, wie bereits erwähnt mit 342 Annotationen, „deutsch“ mit Abstand an erster Stelle. Mit 131 Annotationen liegt „englisch“ weit dahinter. Es folgen auf den Plätzen drei bis zehn „chinesisch“ (35), „amerikanisch“ (24), „französisch“ (23), „arabisch“ (21), „italienisch“ (16), „griechisch“ (15), „preußisch“ (14) und „russisch“ (9). Die quantitativ mit Abstand größte Bedeutung von „deutsch“ und die, im Vergleich zum 19. Jahrhundert, abnehmende Tendenz, andere Nationalitäten zu erwähnen, deutet darauf hin, dass das Nationalempfinden und Deutschtum, das im 20. Jahrhundert zu den beiden Weltkriegen führte, sich auch in der literarischen Raumdarstellung niederschlägt. Zwar ist auch hier wieder das geographische Feld, dass durch nationalstaatliche Zuschreibungen erwähnt wird, relativ groß (umfasst jetzt auch das fernöstliche China und mit „arabisch“ auch wieder nah-südöstliche Länder), die Annotationszahlen sind aber so gering, dass es sich dabei auch um

[24] Die tabellarische Übersicht der sortierten Raumbeschreibungen nach Häufigkeit im Jahrhundertvergleich befindet sich im Forschungsdaten-Repository unter „Datenbasis der Analyse / Raumbeschreibungen / Uebersicht Raumbeschreibungen sortiert Jahrhundertvergleich“ (Schumacher, 2021a).

Einzeltext- oder Textgruppen-Phänomene handeln kann[25]. Das hier am häufigsten vorkommende länderspezifische Adjektiv „deutsch“ findet sich dagegen in 13 der 25 Romane in diesem Teilkorpus. Dass die vergleichsweise häufige Erwähnung von „deutsch“ tatsächlich ein Spezifikum des dritten Teilkorpus ist, zeigt abschließend auch der Vergleich mit dem vierten Teilkorpus. Hier ist „deutsch“ zwar auch wieder das häufigste Adjektiv, das auf eine Länderzugehörigkeit schließen lässt (die entweder über eine tatsächliche Zugehörigkeit zu einem Land aufgerufen wird oder aber einen Sprachraum betreffen kann, der indirekt aber auch wieder lokale Rückschlüsse erlaubt), allerdings wurde es insgesamt nur 116 Mal annotiert. Auf den Plätzen zwei bis zehn stehen hier „englisch“ (40), „amerikanisch“ (23), „italienisch“ (22) „arabisch“ (20), „griechisch“ (18), „französisch“ (13), „holländisch“ (12), „britisch“ (11), „chinesisch“ (8). Auch hier ist nicht nur eine Abnahme der Annotationen des Adjektivs „deutsch“ zu erkennen, sondern auch insgesamt eine Abnahme der mit Sprachen oder Staaten assoziierten Annotationen anderer Types. Dass in allen vier Korpora englisch immer von großer Bedeutung ist, liegt zum Teil auch daran, dass die englische Sprache als lingua franca im westlich-europäischen Raum so bedeutsam ist.

Mit niedrigeren Annotationszahlen müssen allerdings auch die Phänomene als weniger belastbar eingeordnet werden. Die zur Cross-Validierung durchgeführte Vergleichsanalyse mit einem nur um 1% Erkennungsgenauigkeit abweichendem Tool (bei dem im Trainingsprozess die Metaphern gelöscht wurden anstatt sie so zu klassifizieren als wären sie im eigentlichen Wortsinne gemeint) hat gezeigt, dass die Reihenfolge der ersten Plätze sich nicht verändert, dass es aber bei den unteren Plätzen mit einem leicht abweichend trainierten Tool hohe Schwankungen der Annotationszahlen und auch der Ränge im Gesamtgefüge der Annotationen gibt. Solche Abweichungen müssen mit einberechnet werden, wenn Methoden wie die in dieser Studie angewendete als Preprocessing eingesetzt werden. Sind sie Kernbestandteil der Studie, so können solche Bewegungen durch Analysen mit leicht abweichenden Versionen eines Tools gute Hinweise darauf geben, welche Beobachtungen sich als belastbar ergeben.

[25] Dass z. B. „chinesisch“ so weit vorne in der Rangfolge steht, hat den Grund, dass einer der Romane (Heyking *Tschun*) in China spielt, bei einem anderen, einem Briefroman, befindet sich einer der beiden Briefschreibenden in China (Heyking *Briefe, die ihn nicht erreichten*). Insgesamt sind 31 der 35 Vorkommnisse des Lemmas „chinesisch“ auf diese beiden Romane verteilt. Es kommt alles in allem in fünf der 25 Romane vor.

8.2.4 Spezifika des vierten Teilkorpus (21. Jahrhundert)

Das Teilkorpus des 21. Jahrhunderts weist die meisten Spezifika auf. Das liegt zum Teil daran, dass es zeitlich am weitesten vom zuerst betrachteten Teilkorpus des 18. Jahrhunderts entfernt ist, zum anderen spielt hier aber sicher auch der Faktor geringerer Kanonisierung mit herein. Die Texte des 21. Jahrhunderts sind zum großen Teil zu neu, um bereits zum literarischen Kanon zu gehören. Es handelt sich nur in wenigen Fällen um Werke, die bereits in der Literaturwissenschaft betrachtet wurden. Der geringere Korpus-Bias zeigt sich z. B. darin, dass im Teilkorpus des 21. Jahrhunderts auch Genres vertreten sind, die im Allgemeinen eher zur Unterhaltungsliteratur gezählt werden, wie z. B. Kriminalromane. Dadurch könnte es in diesem Teilkorpus zu einer größeren Vielfalt kommen, die sich auch auf die literarische Raumdarstellung auswirken kann.

Insgesamt sind vier Aspekte bei der Auswertung der Annotationsdaten aufgefallen: In diesem Teilkorpus werden erstens besonders viele unterschiedliche Heterotopien erwähnt. Zweitens ist die Leere ein besonders häufig auftauchendes Raumthema, dass sich auch in den Raumbeschreibungen niederschlägt. Inhaltlich sehr ähnlich wie das Raumthema der Leere sind drittens im Teilkorpus des 21. Jahrhunderts die relativ häufig vorkommenden Raumbeschreibungen der Fremde ausgestaltet. Viertens gibt es eine Auffälligkeit im Hinblick auf die Beschreibung von Figuren, die hier vergleichsweise oft als dünn oder dick klassifiziert werden. Diese vier Besonderheiten des literarischen Raumes im 21. Jahrhundert – die Heterotopien, die Themen der Leere und Fremde und die Klassifizierung von Figuren als „dick“ oder „dünn“ – werden im Folgenden kurz beleuchtet.

Darstellung von Heterotopien im Teilkorpus des 21. Jahrhunderts
Der Spiegel, der nach Foucault ein Zwischenraum von Utopie und Heterotopie ist (vgl. Foucault, 2005, 321 und Abschnitt 2.2.2 dieses Buches), steigt langsam über die Jahrhunderte auf eine immer höhere Position innerhalb des Gefüges der Raumthemen. Waren Spiegel im 18. Jahrhundert mit 144 Annotationen noch auf Platz zehn der Raumthemen, so stehen sie mit 139 Annotationen im 19. Jahrhundert auf Platz neun, im 20. Jahrhundert mit 115 Annotationen auf Platz sieben und mit 148 Annotationen im 21. Jahrhundert auf Platz sechs. Es gibt im 21. Jahrhundert aber nicht nur mehr Spiegel in der Literatur als zuvor, auch die heterotop organisierten Raumthemen der Bestattung und Aufführung (im theatralen Sinne) werden hier häufiger thematisiert als in den Jahrhunderten davor (eine Ausnahme bildet das 18. Jahrhundert, in dem sowohl die Thematik der Bestattung als auch der Aufführung auf ähnlichen Plätzen im Gefüge der Raumthemen stehen). Von 64 (19. Jahrhundert) über 45 (20. Jahrhundert) zu 70 Annotationen ist zwar keine große Veränderung, die Tendenz trägt aber zur allgemein häufigeren Erwähnung von Heterotopien bei.

Bei Aufführungen ist interessant, dass sie zwar im 18. Jahrhundert relativ häufig (97 Annotationen) erwähnt werden, danach aber sehr viel seltener raumthematisch verarbeitet sind (18 Annotationen im 19., 15 im 20. Jahrhundert). Mit 49 Annotationen wird diese Heterotopie in der Gegenwartsliteratur wieder etwas bedeutsamer. Auch das ist nur ein leichter Trend, der aber dennoch das heterotope Gesamtarrangement stützt. Dass mehr heterotope Orte als Raumthemen annotiert wurden, zeigt, dass die Darstellung von Abweichungen hier an Bedeutung zunimmt. Ob Tod, Mienenspiel oder auch der Spiegel als realweltlicher Blick in eine utopische nicht-Welt, hier werden Räume thematisiert, die gesellschaftliche und individualpsychologische Konflikte aufgreifen, organisieren oder auf andere Weise eine Auseinandersetzung damit erlauben. Wenn Literatur heterotoper wird, so trägt sie auch selbst dazu bei, diese Art von Themen zu verarbeiten. Der literarische Raum bekommt hier also eine leicht heterotope Funktion, die quantitativ zwar nicht überwältigend ist, aber doch in den Annotationen der Raumthemen sichtbar wird.

Das Raumthema der Leere in der Literatur des 21. Jahrhunderts
Das Raumthema der Leere taucht in allen vier Teilkorpora als solches auf und klettert über die Jahrhunderte langsam nach oben in der Rangliste der häufigsten Raumthemen. Mit 36 Annotationen ist es im 18. Jahrhundert auf Platz 16, mit 16 im 19. Jahrhundert auf Platz 19. Im Teilkorpus des 20. Jahrhunderts kommt es mit 50 Annotationen auf den 12., im zeitgenössischen Teilkorpus mit 84 Annotationen auf Platz 10. Das ist zwar eine relative Zunahme der Bedeutung innerhalb der Raumthemen der Teilkorpora, aber die absoluten Zahlen sind nicht besonders beeindruckend. Allerdings haben wir hier einen potentiellen semantischen Unterschied zwischen dem Nomen „Leere“ und dem Adjektiv „leer“, der dazu führt, dass das Modell ersteres als Raumthema, letzteres als Raumbeschreibung erkennt. Der Raum-Classifier differenziert hingegen nicht zwischen einer raumthematischen Verwendung des Adjektivs „leer“ (denkbar wäre eine Wendung wie „der Raum wirkte leer auf sie, obwohl er es gar nicht war“) und einem raumbeschreibenden Gebrauch („die Tasse war leer“). Während der Analyse der Annotationsdaten hat sich gezeigt, dass hier ein Blick auf die Raumbeschreibungen lohnt. Sowohl auffallend als auch für das Teilkorpus des 21. Jahrhunderts typisch ist, dass hier „leer“ auf Platz 3 der am häufigsten als Raumbeschreibung annotierten Wörter steht. Ebenfalls bezeichnend ist der Vergleich mit der Anzahl der Annotationen der Wörter, die sich auf das Lemma „leer“ zurückführen lassen. Im Teilkorpus des 18. Jahrhunderts betrifft das 230 Annotationen, im 19. Jahrhundert wurde 251 Mal „leer“ annotiert, im 20. Jahrhundert 265 und im 21. Jahrhundert schließlich mit 481 Annotationen beinahe doppelt so viel. Das liegt maßgeblich daran, dass die raumthematische Verwendung des Wortes „leer“ im 21. Jahrhundert im Vergleich zum 20. Jahrhundert zunimmt. Das Adjektiv „leer“ scheint immer dann von einer Raumbeschreibung zu einem

Indikator für ein Raumthema zu werden, wenn es sich auf die Figuren bezieht. Das kann der Fall sein, wenn es entweder um innere Leere geht oder darum, dass der Raum um eine oder mehrere Figuren herum leer ist. Innere Leere ist vorhanden, wenn Figuren direkt als leer beschrieben werden (z. B. Zaimoglu *Isabel* „sie hatte sich leer geredet") oder die für die Innenwelt oder deren Grenze zur Umgebung stehenden Körperteile Kopf, Herz und Augen entsprechend dargestellt werden (z. B. in Berg *Vielen Dank für das Leben* „Tom schaute leer" oder in Mora *Das Ungeheuer* „Mein Herz war leer" oder in Berg *Vielen Dank für das Leben* „Trotz im Blick, abgeprallt an leeren Augen"). Dass die äußere Leere häufig als beklemmend wahrgenommen wird, zeigt die nicht selten vorkommende Verknüpfung mit weiteren Adjektiven, wie z. B. in „wüst und leer war dieses Areal am Anhalter" in Lange Müllers *Böse Schafe*. Zum Teil ist dieses Raumthema auch mit einem anderen verknüpft, nämlich mit dem der Einsamkeit (z. B. in der Passage „plötzlich empfindet sie, wie leer das Haus ist" in Geiger *Es geht uns gut* oder in „das Bett neben mir ist leer, die Wohnung still wie ein Grab" in Moosdorf *Die Freundinnen*). Doch auch wenn die beiden Themen eng miteinander verbunden sein können, wurden sie hier als zwei Einzelthemen beibehalten, denn die Leere steht häufig auch für eine gewisse Emotionslosigkeit oder aber eine negative emotionale Färbung, während Einsamkeit sowohl in negativer als auch in positiver Weise sehr emotional sein kann. Dass die Leere hier so eine Bedeutung erlangt, und das als emotionaler Zustand und damit eher als innenweltliches literarisches Raumthema, unterstützt die Tendenz des zunehmenden Erzählens von Privatem. Die Gefühlswelten der Figuren rücken in den Mittelpunkt. Dass diese emotionalen Räume nicht erfüllt, sondern eben leer sind, kann auf einen gewissen Zeitgeist hindeuten, in dem trotz oder gerade aufgrund von äußerer Fülle wie sie z. B. Globalisierungs- und Digitalisierungsphänomene mit sich bringen, keine innere Zufriedenheit erlangt werden kann.

Raumbeschreibungen der Fremde in der Literatur des 21. Jahrhunderts
Die Annotationen in der Kategorie der Raumbeschreibungen, die Worte umfassen, die sich auf das Lemma „fremd" zurückführen lassen, zeigen zwar insgesamt eine abnehmende Tendenz vom 18. bis zum 21. Jahrhundert (mit einem leichten Peak im 19. Jahrhundert), aber es gibt hier zwei durchaus interessante und überraschende Phänomene. Im 20. Jahrhundert zeigt sich, dass eine Mehrzahl der insgesamt 369 Annotationen in einem einzigen Text vorkommen. Dieses Phänomen wird in Abschnitt 8.3 genauer beschrieben. Im 21. Jahrhundert zeigt sich hingegen eine inhaltliche Verschiebung. Hier wird zunehmend eine emotionale Entfremdung als „fremd" bezeichnet.[26]

---

[26] Die Daten zu diesem Kapitel liegen im Ordner „Datenbasis der Analyse / Raumbeschreibung / Fremdheit" im GitHub-Repository zu dieser Studie (Schumacher, 2021a). Es handelt sich um Tabellen, die die Kontexte der Ausdrücke des Lemmas „fremd" in den vier Teilkorpora

Im Teilkorpus des 18. Jahrhunderts wird das Fremde entweder mit der Reisethematik, mit Figuren, einzelnen Körperteilen oder auch Besitzverhältnissen verknüpft. Auch in Metaphern kommt das Wort „fremd“ vor. Typisch für die Verbindung mit dem Raumthema der Reise sind Phrasen wie „fremde Gegend“ (z. B. in Wieland *Geschichte des Agathon*) oder „fremde Himmel“ (z. B. in Hölderlin *Hyperion oder der Eremit in Griechenland*). Figuren werden z. B. wie folgt beschrieben: „die fremde Frau da“ in Unger *Julchen Grünthal* oder „der fremde Fürst“ in Heinse *Hildegard von Hohenthal*. In *Andreas Hartknopf* von Moritz zeigt eine der Hauptfiguren genannt Knapp „über zwei fremde Gräber weg“; auf diese Weise wird hier das nicht-Eigentum ausgedrückt. Auffallend häufig sind Kombinationen von „fremd“ und Bezeichnungen einzelner Körperteile von Figuren, wie z. B. fremde Hand, fremdes Herz, fremder Kopf. Zu dieser Art Phrasen gehört auch die räumliche Metapher „in fremde Hände fallen / geraten“ bzw. „in fremden Händen sein“ für entführt oder gefangen werden. Bei diesem Ausdruck fällt allerdings auf, dass die gesamte Phrase zwar metaphorisch gebraucht wird, dass die Raumbeschreibung „fremd“ hier aber trotzdem auch als solche wirkt, da sie das nicht-Eigene und häufig auch räumlich andere impliziert. Der Ausdruck wirkt innerhalb der metaphorischen Phrase im Prinzip auf ähnliche Weise vieldeutig wie in einem nicht-metaphorischen Ausdruck (eine Tatsache, die an dieser Stelle noch einmal bestätigt, dass es sinnvoll sein kann, wie hier geschehen, räumliche Metaphern im Trainingsprozess in die übrigen Raumkategorien zu integrieren, statt sie als eigene Kategorie zu betrachten). Ein nahezu identisches Bild ergibt sich für das Teilkorpus des 19. Jahrhunderts. Auch hier werden meist fremde Gegenden, Figuren, auch mal Kleidungsstücke oder Pflanzen bezeichnet. Nahezu immer wird entweder ein Subjekt oder ein Objekt als fremd bezeichnet, seltener wird eine immaterielle Größe wie Macht als fremd bezeichnet. Emotionale Fremdheit spielt hier noch keine Rolle. Auch im Teilkorpus des 20. Jahrhunderts ist „fremd“ eine sehr häufige Raumbeschreibung. Mit 369 Annotationen steht es hier auf Platz fünf. Die inhaltlichen Konnotationen unterscheiden sich nicht sehr von denen des 18. und 19. Jahrhunderts. Die Vielfalt an immateriellen Größen, die mit diesem Ausdruck bezeichnet werden, nimmt etwas zu. Eine „fremde, unbegreifliche monströse Wirklichkeit“ kommt z. B. in Rilkes *Malte* vor. In Keyserlings *Abendliche Häuser* sieht eine Figur immer wieder das Gesicht einer anderen vor sich „mit den flackernden Augen und dem fremden Ausdrucke von Erregung und Qual“.

Ähnlich wie im Teilkorpus des 20. Jahrhunderts ist auch im vierten Teilkorpus das Wort „fremd“ relativ weit oben bei den am häufigsten annotierten Raumbeschrei-

---

zeigen und die Daten, die durch das Topic Modelling erstellt wurden. Letztere befinden sich im Unterordner „topicsexplorerTeilkorpus20 15Topics 5000Iterationen“.

bungen dabei (Platz zehn). Die absolute Anzahl (209 Annotationen) und auch die Platzierung hat jedoch gegenüber den anderen drei Teilkorpora deutlich abgenommen. Und auch wenn hier ebenfalls wieder fremde Figuren (auch Blicke, Augen, Hände u. Ä.) und Gegenden oder auch Sprachen vorkommen, so fällt doch eine Kontextverschiebung auf. Häufig geht mit räumlicher Fremde auch emotionale Fremdheit oder Entfremdung einher. Figuren wirken auf ein Mal fremder („Der eigene Mann wird ihr immer fremder" in Gruber *Über Nacht*) oder eine Figur, die einer anderen nahe stehen sollte, bleibt fremd („Mein krankes, fremdes, anderes Kind" in Steinfests *Allesforscher*). Ebenfalls häufig sind fremde Straßen, Städte oder Welten. Wie zuvor schon, fällt auch hier wieder das Zusammenspiel von Innen- und Außenwelten auf. Das Gefühl, fremd zu sein, nicht dazuzugehören, wird aus einem räumlichen Kontext auf einen emotionalen übertragen. In der Passage „Monika. Minka. Minoka. Gleichzeitig vertraut und trotzdem so fremd, fremder, ferner Stern." aus Weins *Lazyboy* wird genau diese Übertragung des Empfindens emotionaler Fremde auf das räumlicher Fremde beschrieben. Auch in der Formulierung „Und wirklich denkt sie jetzt, während sie auf dem Marktplatz eines fremden Dorfes Münzen einsammelt – wäre sie noch einmal in die Bäckerei gegangen, so wäre sie geblieben und hätte den Steger-Sohn geheiratet, den älteren, dem vorne zwei Zähne fehlen." aus Kehlmanns *Tyll* steht die räumliche Fremde für das Ungewohnte, die Abweichung vom vorgezeichneten Lebensweg. Dass hier viel mehrdeutigere Implikationen als bloße Raumbeschreibungen vorliegen, machen die Beispiele sehr deutlich. Hierin liegt auch ein Grund dafür, dass die Annotationen der Raumbeschreibungen nur zu 60% in die Berechnung des Raumindexwertes (RIW) einfließen. Das Beispiel der Fremde zeigt darüber hinaus, dass die implizite Räumlichkeit von Beschreibungen über die Jahrhunderte nicht konstant sein muss, sondern dynamisch zu- oder abnehmen kann.

Dick oder dünn – Körperklassifizierungen und Schutzzonen in der Literatur des 21. Jahrhunderts

Eine weitere in diesem Teilkorpus relativ häufige Raumbeschreibung ist „dick" (bzw. Variationen des Wortstamms) auf Platz fünf. Diese Häufung liegt zum Teil daran, dass in Kehlmanns *Tyll* eine Figur („der dicke Graf") sehr häufig mit diesem Adjektiv aufgerufen wird. Genau 122 der insgesamt 408 Annotationen des Wortes „dick" finden sich in diesem Roman. Die übrigen sind allerdings über das ganze Teilkorpus verteilt. In allen 25 Romanen wurde mindestens einmal dieses Wort als Raumbeschreibung annotiert. Es handelt sich also nicht um ein Einzeltextphänomen. Neben Figuren, die sehr häufig als dick beschrieben werden, wird der Ausdruck auch verwendet, um Schutz und / oder Komfort bis hin zum Überfluss zu implizieren. Dicke Decken, Kissen, Pullover und Mäntel kommen ebenso

vor wie dicke (schützende) Glasscheiben und Eisdecken. Eine eingehende Prüfung hat ergeben, dass insgesamt 235 Mal in diesem Teilkorpus mit dem als Raumbeschreibung annotierten Wort „dick“ eine Figur beschrieben wurde.[27] Allein 105 Mal geschieht das in Kehlmanns *Tyll*. Für die anderen Romane bleiben also noch 130 Vorkommnisse. Davon werden 31 Mal einzelne Körperteile als „dick“ beschrieben. Mit insgesamt 251 annotierten Wörtern, die zu den Lemmata „dünn“ und „dürr“ gehören, sind diese Pendants zu dick ebenfalls häufig („dünn“ ist in diesem Teilkorpus auf Platz neun der am häufigsten annotierten Wörter). Betrachtet man auch diese Annotationen genauer, so ergibt sich ein überraschend ausgeglichenes Bild. Bei 90 der Annotationen des Ausdrucks „dünn“ handelt es sich um Beschreibungen einer Figur. Davon spezifizieren 25 einzelne Körperteile. Die Klassifizierung von Körpern, Körperteilen und Objekten in „dünn“ und „dick“ nimmt demnach nicht nur an Bedeutung zu. Es scheint im 21. Jahrhundert sogar einen Trend zur Binärklassifikation in diese Körperkategorien zu geben, bei der eine relative Ausgeglichenheit herrscht und „dicke“ Figuren nur ganz leicht überwiegen (von überdurchschnittlichem Gebrauch des Wortes „dick“ als Charakterisierung einer bestimmten Figur in einem einzelnen Roman einmal abgesehen).

Dieser Ansatz einer an quantitativen Beobachtungen orientierten Datenanalyse könnte nun noch weiter verfolgt werden, um genauere Körperbilder der Literatur des 21. Jahrhunderts zu beschreiben. Das ist aber nicht das Ziel dieser Studie, die anhand kurzer Beispielanalysen eher zeigen soll, dass die entwickelte und angewendete Methodik solche Einstiegspunkte in die Analyse bestimmter literarischer Motive und Ausgestaltungen bieten kann. Die Fülle der erhobenen Daten führt nicht unbedingt zu sehr genauen Interpretationen. Es zeigen sich aber viele Besonderheiten und Trends, die dann weiter untersucht werden können. Dazu bieten sich die hier beispielhaft gezeigten Unterklassifizierungen der Raum-Annotationskategorien oder auch die Betrachtung von einzelnen auffälligen Wörtern oder Wortgruppen an.

8.3 Perspektive: Einzeltexte

Unabhängig davon, ob digitale Methoden dazu genutzt werden, an analoge Ansätze anzuknüpfen oder ob ein explorativer Zugang gewählt wird, stellt sich eine Frage, die stets im Blick behalten werden sollte: Sind eventuelle Verzerrungen der Datenbasis durch dominante Einzeltextphänomene vorhanden? Auch im Kernkorpus die-

[27] Tabellen, in denen die Ausdrücke „dick“ und „dünn“ im Hinblick darauf klassifiziert wurden, ob sie Figuren oder Objekte beschreiben, befinden sich im GitHub-Repository zu dieser Arbeit unter „Datenbasis der Analyse / Raumbeschreibung / duenn dick“ (Schumacher, 2021a).

ser Studie ist ein solches Einzeltextphänomen aufgefallen, das in diesem Kapitel betrachtet wird, um zu zeigen, inwiefern solche Ausreißer die Analyse verzerren können. Der Schritt zurück von den Annotationen als „Raumbeschreibung" in einem Teilkorpus zur thematischen Verankerung und auch Ausrichtung in einzelnen Texten wird hier mit Hilfe der Methodik des Topic Modeling vorgenommen. Die Ergebnisse dieser ergänzenden digitalen Methode dienen der Cross-Validierung der bei der Datenanalyse der Annotationen des Raum-Classifiers aufgestellten Thesen. Der betrachtete Beispielfall betrifft die Häufung der Raumbeschreibung „fremd" in einem bedrohlich-negativen Kontext, die bei der Analyse des Teilkorpus des 20. Jahrhunderts zu Tage kam. Topic Modeling hat den Vorteil, dass durch die Gruppierung anderer Ausdrücke, die häufig mit Wörtern, die Fremde ausdrücken, zusammenstehen, nicht nur deutlich wird, dass ein Thema für ein Dokument von Bedeutung ist. Über die Zusammensetzung der Topics tritt außerdem hervor, welche Ausrichtung das Thema im Text hat. Die fokussierten Einzeltexte sind Heykings *Tschun* und Mays *Im Lande des Mahdi I und II*, betrachtet wird die thematische Konnotation des Raumthemas der Fremde.

Bedrohung vs. Exotik – Fremdheit in Heykings *Tschun* und Mays *Im Lande des Mahdi I und II*

Insgesamt wurden im Teilkorpus des 20. Jahrhunderts 369 Vorkommnisse des Wortes „fremd" als Raumbeschreibung annotiert. Fast 1/3 davon (genauer 133 Annotationen) kommen in zwei Romanen von ein und derselben Autorin vor, nämlich in Heykings *Tschun* und *Briefe, die ihn nicht erreichten*. Bei der genaueren Betrachtung fallen zwei Dinge besonders auf. Im Vergleich zum Teilkorpus des 19. Jahrhunderts sind Wendungen wie „fremde Truppen", „fremde Soldaten" oder auch „fremde Teufel" häufig. Diese kommen, von wenigen Ausnahmen abgesehen, nur in den bereits benannten Romanen von Heyking vor. Hier scheint das Thema der (bedrohlichen) Fremdheit eine Rolle zu spielen. Von den 369 Annotationen von „fremd" stammen allein 110 aus dem Roman *Tschun*, weitere 23 aus *Briefe, die ihn nie erreichten*. Dass „fremd" hier ein besonders häufiges Wort ist, das meist in engem Zusammenhang mit Wörtern wie China, chinesisch, Tschun, Kaiser, Kaiserin oder Peking steht, zeigt ein Topic-Modeling[28] des Teilkorpus (vgl. Abbildung 8.14).

---

[28] Dafür wurde der DARIAH TopicsExplorer (Simmler et al., 2019) genutzt. Topic Modeling ist eine Methode, die häufig zusammenstehende Wörter eines Korpus gruppiert und die Verteilung dieser Gruppen in den Dokumenten des Korpus anzeigt. Topic Modeling ist in dieser Arbeit keine zentrale Methode und wird lediglich zur Cross-Validierung der aufgestellten Thesen verwendet. Eine ausführliche Methodenreflexion würde darum im Rahmen dieser Arbeit zu weit führen.

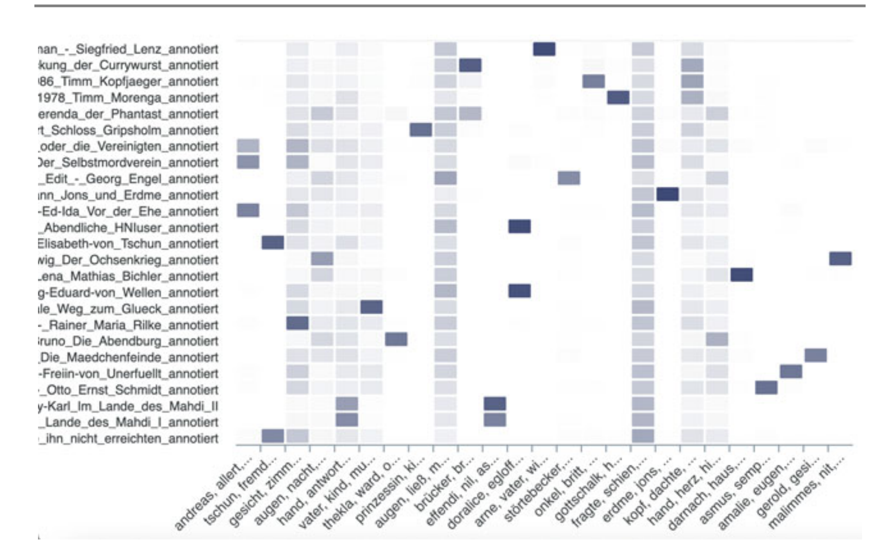


Abbildung 8.14 Topic Modeling des Teilkorpus 20. Jahrhundert: Das Topic „Tschun, fremd, . . .“ ist ausschließlich für Heykings Romane von Bedeutung.

Sowohl „fremden“ als auch „fremde“ sind in einem Topic relativ weit vorne in der Wortliste, nämlich auf den Plätzen 2 und 57, was bedeutet, dass diese Wörter besonders häufig im Zusammenhang mit den anderen in dem Topic häufigsten Wörtern stehen. Diese Analyse zeigt auch, dass dieses Topic in den beiden bereits genannten Romanen von Heyking besonders präsent ist. Da Fremdheit für die anderen Topics des Korpus weniger bedeutsam ist (es kommt in drei weiteren Topics unter den ersten 100 Wörtern vor, allerdings nie mehr unter den ersten 50), scheint es sich hierbei tatsächlich um ein Ausreißer-Phänomen zu handeln. Die Thematik mag für die Romane Heykings zwar von großer Bedeutung sein, eine Verallgemeinerung auf das gesamte Teilkorpus oder gar die Annahme eines Trends der Literatur des 19. Jahrhunderts ist auf Basis dieser Daten aber vermessen. Bei einem genauen Blick auf das Topic Modeling des Korpus’ mag die Frage aufkommen, ob das Thema der Fremdheit nicht auch in den beiden Romanen von Karl May *Im Lande des Mahdi I und II* bedeutsam ist. Tatsächlich zeigt das Topic Model, dass es ein Karl-May-spezifisches Topic gibt (analog zum Heyking-Topic). In diesem Topic kommt aber kein Begriff vor, der Fremdheit explizit ausdrückt. Stattdessen sind hier ungewöhnliche und exotisch anmutende Namen besonders häufig. Auch „Kamele“ und „Wüste“ sind unter den ersten 100 Wörtern dieses Topics. Statt der Fremde bei

Heyking haben wird hier eher Exotik als Thema etabliert. Der Sprung zurück in die Annotationsdaten in CATMA zeigt, dass das Wort „fremd“ zwar auch in Karl Mays Romanen vorkommt, aber immer zur Bezeichnung des „fremden Effendi“. Auch hiermit wird also Exotik erzeugt und weniger Bedrohlichkeit.

Die beispielhaft angeführten Analysen und Interpretationsansätze haben zwei Dinge deutlich gemacht. Es ist erstens zwar möglich mit Hilfe des Raum-Classifiers Einstiegspunkte in die Interpretation literarischen Raumes zu finden. Die Fülle an automatisch generierten Annotationen macht es zweitens jedoch schwer, allein auf Basis des sechs Kategorien umfassenden Raummodells ausführliche und abschließende Beschreibungen des literarischen Raumes im gesamten Kernkorpus zu erstellen. Häufig ist es Gewinn bringend, die Annotationen einer Kategorie weiter zusammen zu fassen und Unterkategorien zu finden. Die automatische Annotation mit Hilfe des Raum-Classifiers hat sich aber als eine Möglichkeit des Preprocessing für ein mittelgroßes, diachrones Kernkorpus erwiesen, die auf unterschiedlichen Ebenen Ansatzpunkte für die Analyse literarischen Raumes bietet. Auf Ebene des gesamten Korpus wurde sichtbar, dass der literarische Raum überwiegend relational dargestellt wird, dass die Verknüpfung von Innen- und Außenwelten große Bedeutung hat und dass ein Spannungsfeld zwischen dem Hier und dem Dort aufgebaut wird. Diachrone Entwicklungen, wie z. B. die zunehmende Schilderung von Privatheit, zeigten sich bei der vergleichenden Analyse der vier Teilkorpora. Als etwas weniger fruchtbar, aber auch möglich, hat sich der Versuch erwiesen, jahrhundertspezifische Phänomene in einem explorativen Zugang auszumachen. Nicht nur die Tendenz des mit Hilfe von Machine Learning trainierten Raum-Classifiers eher übergreifende Phänomene als neu auftretende zu erkennen, erschwert einen solchen Ansatz. Um Spezifika zu finden, muss immer erst eine Ausgangsmenge definiert werden. Bei einer diachronen Analyse wie der vorliegenden, führt die Perspektive dazu, dass die zuerst betrachteten Korpora als eine solche Norm-Menge wahrgenommen werden. Es ist perspektivbedingt schwieriger, darin Besonderheiten zu finden. Das zeigt sich auch in den in Abschnitt 8.2 angeführten Beispielen und zwar dadurch, dass für das Teilkorpus des 21. Jahrhunderts die meisten Besonderheiten ausgemacht werden konnten. Liegt der Fokus auf quantitativ heraustretenden Merkmalen eines Korpus, so ist es eher selten, dass Einzeltextphänomene zu Tage kommen. Dass dies aber mit der hier entwickelten Methodik auch passieren kann und darum auch berücksichtigt werden sollte, zeigt das als letztes angeführte Beispiel der Fremdheit in Heykings *Tschun* und *Briefe die ihn nicht erreichten*. Bei dieser letzten Beispielanalyse hat es sich als besonders gewinnbringend erwiesen, eine weitere digitale Methode zur Cross-Validierung zu nutzen. Mit Hilfe des Topic Modeling konnte nicht nur belegt werden, dass das Fremdheits-Topic sich auf Heykings Romane beschränkt, sondern auch, dass es ein zwar ähnliches aber gegentei-

lig, nämlich positiv, konnotiertes Topic in Mays *Im Lande des Mahdi I und II* gibt. Es gäbe noch eine Fülle von Phänomenen, die hier angeführt werden könnten, da sie sich in den Annotationsdaten finden. Da die hier beschriebenen Beispiele aber bereits sehr gut als Proof of Concepts dienen und zeigen, dass die Vorverarbeitung mit Hilfe des Raum-Classifiers auf allen drei Ebenen – der des gesamten Korpus, der der einzelnen Teilkorpora und der des Einzeltextes – für die Untersuchung literarischer Raumphänomene eingesetzt werden kann, wird hier darauf verzichtet, bloß der Vollständigkeit halber, weitere Beispiele anzuführen, die keine wesentlichen neuen Aspekte der Methodik aufzeigen.

9 Diachrone Entwicklung der Bedeutung von Raum in literarischen Texten

Ausgangspunkt dieser Arbeit war die Frage, wie die Darstellung von Raum in literarischen Texten in einem quantitativen Ansatz untersucht werden könnte. Bisher wurde ein Teilaspekt dieser Fragestellung beantwortet: Mit Hilfe eines theoriebasierten Raum-Classifiers zur automatischen (Vor-)Annotation literarischer Texte können Einstiegspunkte in die Interpretation literarischer Raumdarstellung gefunden werden und es können übergeordnete Perspektiven auf das gesamte Kernkorpus, die Teilkorpora oder Einzeltexte eröffnet werden. Nun soll eine weitere Möglichkeit, Raumdarstellung in literarischen Texten untersuch- und messbar zu machen, betrachtet werden. Im Folgenden ist dokumentiert, wie der Raum-Classifier, bzw. dessen automatisierte Annotationen, dazu dienen können, einen Raum-Index-Wert (RIW) pro Text zu ermitteln, um den Anteil der Raumdarstellungen der Romane miteinander vergleichen zu können. Auch hier wird eine diachrone Perspektive eröffnet. Wird das gesamte, im Raum-Classifier implementierte, Raummodell angewendet, so wird das Ergebnis zeigen, wie hoch der Anteil der Darstellung von Raum in Texten ist, wenn ein relativ weit gefasster Raumbegriff zu Grunde gelegt wird. Nun kann das in dieser Studie entwickelte theoriebasierte fuzzy-set-Modell aber durchaus variabel genutzt werden. Es können alle sechs Raumkategorien zu einem weiten Raumbegriff zusammengefasst werden, wie bei der Ermittlung des RIW pro Text der Fall. Es können aber auch Unterkategorien zusammengefügt werden. Die Beispielanalyse in Abschnitt 8.1.1 mündete in die These, dass der Raum, der in literarischen Texten dargestellt wird, hauptsächlich relational ist. Soll nun für den relationalen Raum ebenfalls ein Index erstellt werden, so können die beiden relational angelegten Kategorien „Relation“ und „relationale Verben“ zusammengefasst werden. Wie auch beim RIW werden beim Relationen-Index-Wert (RelIW) die Relationen zu 100 % mit einberechnet, die relationalen Verben aber nur zu 80 %, da diese häufig mehr als nur Rauminformationen vermitteln. Eine Einzelkategorie, die sich eignet, um einen Indexwert für einen engen Raumbegriff zu errechnen, ist

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_9

die der Orte. Darum wurden hier pro Text auch noch je ein Orts-Index-Wert (OIW) errechnet, der den Raum-Anteil zeigt, den die Romane erreichen, wenn ein enger Raum-Begriff angenommen wird.[1]

Mit diesen drei Werten kann die gesamte Darstellung von Raum der Erzähltexte, die unterliegende Netzwerkstruktur räumlicher Relationen und das Setting erfasst werden. In der Literaturwissenschaft wurde aber auch herausgearbeitet, dass (implizite) Rauminformationen auf thematischer Ebene literarisch bedeutsam sind. Darum werden auch die Anteile von Raumthemen in den Texten mit Hilfe eines Indexwertes (ThIW) berechnet. Da die Wörter, die als Raumthemen zugehörig annotiert wurden oft implizit räumlich sind und noch einige weitere Aspekte beinhalten, werden sie – ebenso wie bei der Berechnung des RIWs – nur zu 70 % in die Rechnung einbezogen. Diese 70 % aller als Raumthema annotierten Wörter werden dann durch die Gesamtwortzahl eines Textes gerechnet.

Auf diese Weise wird gezeigt, dass das Modell genutzt werden kann, um anteilig Raumdarstellung zu berechnen für:

1. Einen weiten und komplexen Raumbegriff.
2. Einen relationalen Raumbegriff, der sich aus zwei Unterkategorien des fuzzy-set-Raummodells ergibt.
3. Einen engen Raumbegriff, der hauptsächlich auf das Setting von Romanen ausgerichtet ist und explizite Ortsnennungen als Indikatoren annimmt.
4. Raumthemen, die häufig über implizit räumliche Wörter aufgerufen werden.

Damit werden hier die Kategorien der Raumhinweise und Raumbeschreibungen nur im Zusammenhang mit dem weiten Raumbegriff mit einbezogen. Das liegt zum einen daran, dass mit weiteren Indexwerten keine neue methodische Variante mehr aufgezeigt werden kann. Zum anderen sind diese Kategorien aber auch diejenigen, die in den Tests des Raum-Classifiers mit durchschnittlich 71,7 % und 59,8 % noch die schlechtesten Ergebnisse gezeigt haben. Im Umkehrschluss wären hier also die am wenigsten belastbaren Einsichten zu erwarten. Darum werden für Raumhinweise und Raumbeschreibungen keine weiteren Indexwerte eingeführt.

Betrachtet man zunächst die Raum-Index-Werte (RIWs) im gesamten Kernkorpus, so wird eine diachrone Entwicklung deutlich. Die RIWs der Texte steigen minimal an (vgl. Abbildung 9.1).

[1] Die Berechnungen der Raum-Index-Werte pro Teilkorpus und Kernkorpus sowie eine Berechnung der Index-Werte anhand der manuellen Annotationen der Testtexte sind im GitHub-Repository zu dieser Arbeit unter „Datenbasis der Analyse / Raum-Index-Werte" zu finden (Schumacher, 2021a).

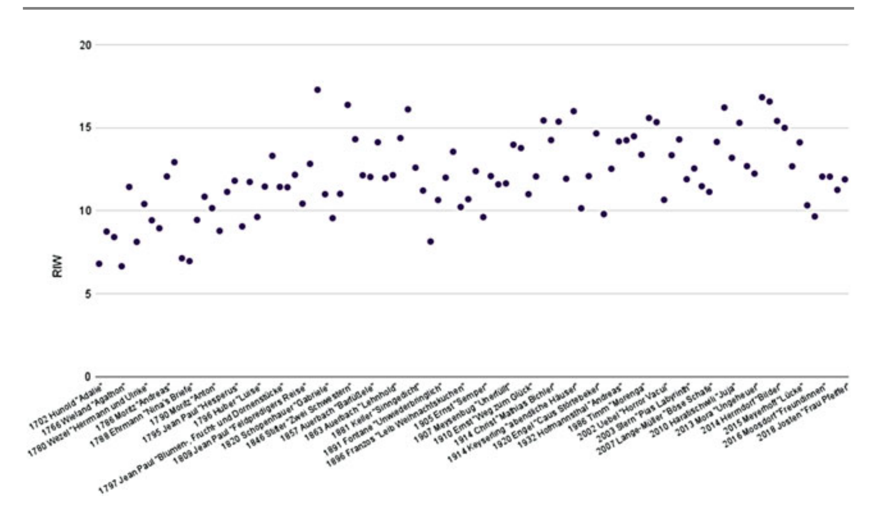


Abbildung 9.1 Raum-Index-Werte (RIWs) im gesamten Kernkorpus

Die Mittelwerte der einzelnen Teilkorpora bestätigen diesen ersten Eindruck. Im Teilkorpus des 18. Jahrhunderts liegt der RIW durchschnittlich bei 9,88. Das heißt, dass im Schnitt rund 10 % des Wortmaterials der Romane Rauminformationen referenzieren. Die Standardabweichung beträgt in diesem Teilkorpus 1,92. Die meisten der Raumwerte bewegen sich also in einem Bereich zwischen 7,96 und 11,8. Im zweiten Teilkorpus liegt der durchschnittliche RIW bei 12,37, der Raumanteil pro Roman beträgt also im Schnitt rund 12 %. Die Standardabweichung ist hier aber mit 2,21 etwas höher. Es gibt also mehr Varianz. Die meisten RIWs liegen hier in einem Bereich zwischen 10,16 und 14,58. Im 20. Jahrhundert ist der RIW-Durchschnitt 13,12. Die Standardabweichung geht vom 19. zum 20. Jahrhundert leicht zurück und liegt nun bei 1,87. Der Normalbereich für RIWs ist also 11,25–14,99. Im 21. Jahrhundert schließlich bleibt der durchschnittliche RIW auf gleichem Niveau bei 13,12, die Standardabweichung ist mit 1,99 ebenfalls sehr ähnlich wie im 20. Jahrhundert. Es entsteht der Eindruck, dass sich der Anteil der Raumdarstellungen über die Jahrhunderte einpendelt. Im 18. Jahrhundert ist der RIW meist vergleichsweise gering, im 19. Jahrhundert beginnt eine Tendenz zu vermehrter Raumdarstellung, aber es gibt hier auch noch mehr Schwankungen. Einige Romane haben hier bereits sehr hohe RIWs, andere bleiben auf dem Stand des 18. Jahrhunderts. Ab Beginn des 20. Jahrhunderts fängt dann eine Konsolidierungsphase an. Heute nimmt Raumdarstellung in der Literatur also buchstäblich mehr Raum ein als noch im 18. Jahrhundert und hat sich als konstante Größe von rund 13 % des Wortmaterials etabliert.

Wie wichtig die relationalen Raumkategorien für dieses Modell sind, zeigt sich, wenn Relationen und relationale Verben zusammen betrachtet werden. Darum gleicht sich auch die diachrone Entwicklung ziemlich stark (vgl. Abbildung 9.2.)[2] Im 18. Jahrhundert liegt der RelIW durchschnittlich bei 6,31 % und die Standardabweichung beträgt hier 1,39. Zum 19. Jahrhundert nehmen die RelIWs durchschnittlich um rund 2 % zu und der Mittelwert erhöht sich auf 8,07. Die Standardabweichung ist mit 1,38 beinahe gleich. Der Anstieg des mittleren RelIWs vom 19. zum 20. Jahrhundert ist dann sehr gering, der Durchschnittswert steigt auf 8,48, die Standardabweichung bleibt mit 1,18 auf einem ähnlichen Niveau, sinkt aber leicht ab. Im 21. Jahrhundert ist der durchschnittliche RelIW dann fast gleich und beträgt 8,28. Die Standardabweichung liegt hier bei 1,64, steigt also wieder leicht an und zeigt eine leicht erhöhte Varianz der Werte. Insgesamt ergibt sich ein relativ konstantes Bild: Räumliche Relationen machen in der Literatur des 18.–21. Jahrhunderts zwischen 6 und 8,5 % des Textes aus. Sie sind ein entscheidender Faktor der literarischen Raumdarstellung und prägen diese maßgeblich. Darum verwundert die starke Ähnlichkeit der Visualisierung der RelIWs zu der der RIWs auch nicht. Auffallend ist hier, dass die Konsolidierung bereits zwischen dem 18. und 19. Jahrhundert einsetzt also etwa 100 Jahre vor der des RIWs. Dies könnte einerseits darauf zurückzuführen

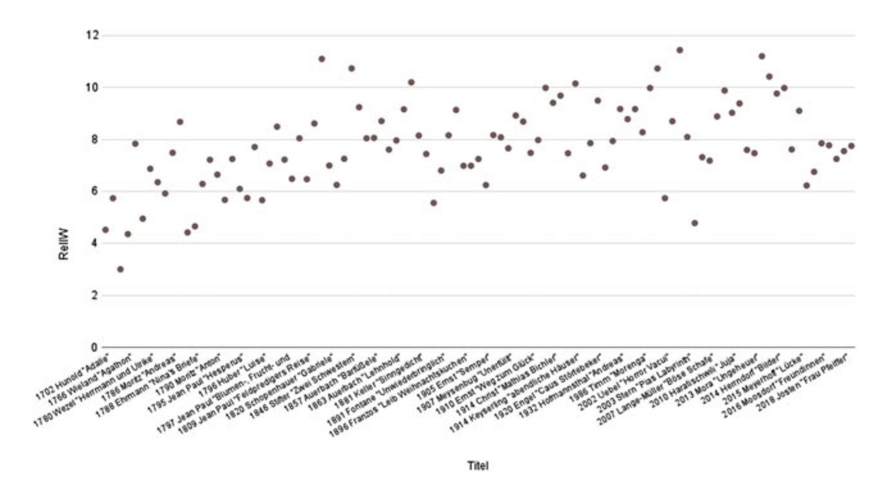


Abbildung 9.2 Relationen-Index-Werte (RelIWs) im gesamten Kernkorpus

---

[2] Ein direkter Vergleich der unterschiedlichen Indexwerte wird am Ende dieses Kapitels anhand von Abbildung 9.5, die Verlaufskurven aller Werte zusammen zeigt, durchgeführt.

sein, dass ein relativer Raumbegriff sich zu dieser Zeit ausgebildet und verfestigt hat. Es zeigt aber auch, dass Relationen eine wichtige Konstante in der Referenzierung von Raum sind und dass diese Konstante für die Literatur über die Jahrhunderte hinweg etwa gleich bedeutsam bleibt.

Ein leicht verändertes Bild ergibt sich, wenn man von einem engen Raumbegriff ausgeht und nur die Ortsannotationen in Relation zur Gesamtlänge der Erzähltexte setzt. Auf diese Weise erhält man einen Orts-Index-Wert pro Text und kann diese miteinander vergleichen. Die diachrone Entwicklung kann wie in Abbildung 9.3 visualisiert werden. Auch hier sind die Werte im 18. Jahrhundert am niedrigsten. Sie liegen zwischen 0,82 und 2,46 % und der Durchschnittswert bei 1,47. Die Standardabweichung beträgt hier 0,38. Im 19. Jahrhundert zeigen sich wieder die größten Schwankungen. Hier liegen die Werte zwischen 0,79 und 3,36 %. Der Mittelwert ist 1,91 und die Standardabweichung beträgt 0,57. Der durchschnittliche OIW steigt also vom 18. zum 19. Jahrhundert an. Im 20. Jahrhundert liegt der OI-Mittelwert bei 1,97, also in etwa auf gleicher Höhe wie im 19. Jahrhundert. Die Standardabweichung ist mit 0,42 etwas niedriger, d.h. die Varianz ist geringer als im 19. Jahrhundert. Tatsächlich liegen die Werte hier zwischen 1,27 und 2,81 %. Im 21. Jahrhundert liegen die Extremwerte bei 1,2 und 2,65. Der Durchschnittswert ist 1,99 also geringfügig höher als im 20. Jahrhundert. Die Standardabweichung sinkt weiter leicht auf 0,36. Wir haben also auch hier einen deutlichen Anstieg vom 18. zum 19.

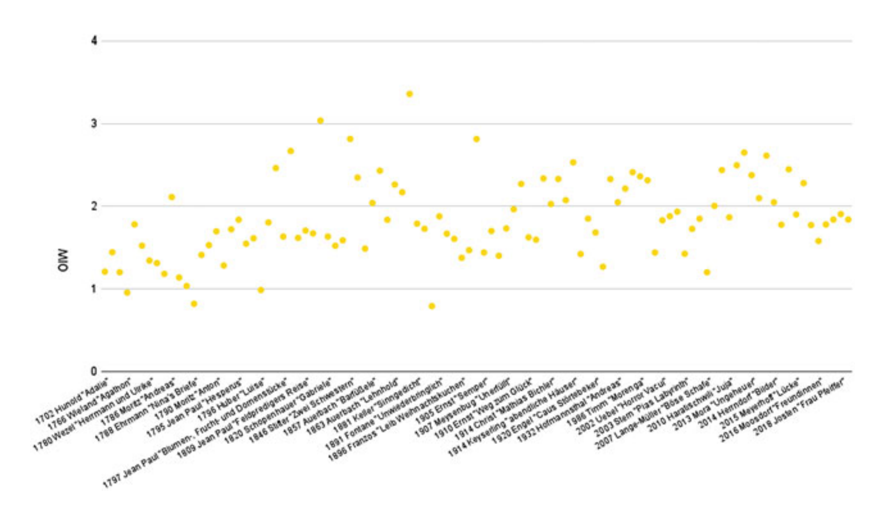


Abbildung 9.3 Orts-Index-Werte (OIWs) im gesamten Kernkorpus

Jahrhundert, dann eine ganz leichte Steigung bis zur Gegenwart. Die Varianz ist im 19. Jahrhundert am höchsten, ebenso wie bei den RIWs. Auch hier setzt die Konsolidierung im 20. Jahrhundert ein und die OIWs haben sich in der zeitgenössischen Literatur auf rund 2 % eingependelt. Das heißt also, dass im Schnitt heute rund 2 % eines Romans aus Referenzierungen von Orten besteht. Anders als bei den RIWs, bei denen Ambiguitäten und Mehrdeutigkeiten mit einbezogen werden, haben wir bei den OIWs eine rein auf Wortvorkommen basierende Berechnung. Von den 100 Texten des Kernkorpus gibt es keinen einzigen, in dem nicht mindestens 0,82 % der Wörter Ortsnennungen sind. Ein Blick auf die absoluten Zahlen der Ortsannotationen zeigt, dass der Text mit den wenigsten Ortsannotationen, Ehrmanns *Nina's Briefe an ihren Geliebten*, immer noch 364 Ortsnennungen bei 46.363 Wörtern Gesamtlänge beinhaltet. Erzähltexte ohne Raumdarstellungen oder auch nur ohne Ortsnennungen kommen in dieser Stichprobe nicht vor.

Nun wurde mit dem gesamten Raum-Kategorien-System ein sehr weiter Raumbegriff betrachtet und mit der Orts-Kategorie ein sehr enger. Gegen Ersteres könnte argumentiert werden, dass der Einbezug von impliziten Raumkategorien, die vielleicht literaturwissenschaftlich gar nicht einheitlich als solche interpretiert würden, den relativ hohen Anteil von Raumdarstellungen in Literatur von durchschnittlich 13 % im gesamten Kernkorpus erklärt. Gegen die Berechnung von Orts-Index-Werten könnte eingewendet werden, dass hier eine Kategorie betrachtet wird, die hauptsächlich zum Setting der Romane beiträgt, also weniger im Vordergrund der Erzähltexte steht. Darum wird nun als letztes noch ein Indexwert für Raumthemen eingeführt.

Die Kategorie der Raumthemen beruht in besonderem Maße auf der literaturwissenschaftlichen Forschung zur Darstellung von Raum, da sie bereits recht umfassend dargestellt und definiert wurde. Gleichzeitig handelt es sich aber um eine Kategorie, die nur zum Teil an Indikatoren im Text festzumachen ist. Objekte, die in der Literaturwissenschaft, z. B. als zur Grenzthematik gehörend, ausgemacht wurden, wie z. B. Wände, Türen und Fenster, tragen nicht nur zu dieser Thematik bei, sondern illustrieren auch das Setting eines Romans. Darum wird, analog zur Berechnung der RIWs, auch hier ein pauschaler prozentualer Anteil von 70 % angenommen, der zeigt, dass hier die Mehrdeutigkeit der Wort-Indikatoren im Text einbezogen wurde. So wurde also für jeden Text im Kernkorpus die Anzahl der zu Raumthemen beitragenden Wörter mal 0,7 gerechnet und das Ergebnis durch die Gesamtwortzahl des Textes dividiert.

Wie Abbildung 9.4 zeigt, verläuft die Entwicklung der Raumthemen im Kernkorpus wieder ähnlich. Insgesamt ist der Anteil von Raumthemen pro Text zwar gering und bewegt sich konstant unter einem Prozent. Aber auch hier ist wieder eine minimal steigende Entwicklung zu erkennen. Im ersten Teilkorpus machen Raumthemen

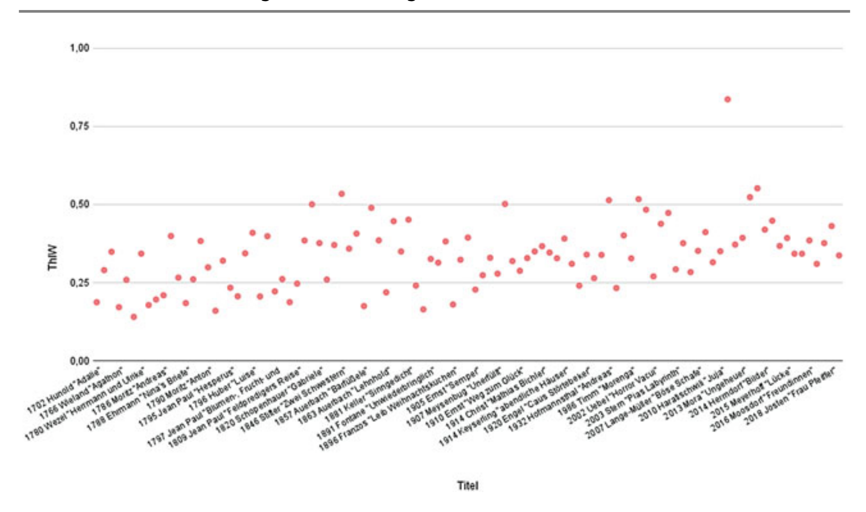


Abbildung 9.4 Raumthemen-Index-Werte (ThIWs) im gesamten Kernkorpus

durchschnittlich 0,27 % eines Romans aus. Die Standardabweichung liegt hier bei 0,08. Die Streuung ist also relativ gering. Im 19. Jahrhundert liegt der Durchschnitt der ThIWs bei 0,34 % und die Standardabweichung bei 0,11. Hier ist der Anteil, den Raumthemen an einem Roman im Schnitt haben, etwas höher als zuvor, aber die Varianz ist auch höher. Es gibt stärkere Schwankungen zwischen dem niedrigsten und dem höchsten Wert. Vom 19. zum 20. Jahrhundert erhöht sich der durchschnittliche ThIW nur leicht auf 0,35. Dafür geht die Standardabweichung hier aber zurück auf 0,08, was zeigt, dass sich die Varianz verringert. Im 21. Jahrhundert liegt der durchschnittliche ThIW schließlich bei 0,4 und die Standardabweichung bei 0,11. Sowohl der Durchschnittswert als auch die Standardabweichung werden im vierten Teilkorpus aber entscheidend dadurch beeinflusst, dass es mit Weins *Lazyboy* einen Ausreißer-Text gibt, in dem Raumthemen 0,84 % des Romans ausmachen. Nähme man diesen Text aus der Rechnung heraus, so läge der durchschnittliche ThIW bei 0,38 und die Standardabweichung bei 0,07. Wieder kann festgestellt werden, dass die Bedeutung von raumbezogenen Thematiken vom 18. Jahrhundert bis zur Gegenwartsliteratur wenn auch nur minimal zunimmt. Es fällt auch auf, dass es selbst bei dieser impliziteren Raumkategorie keinen Text mit einem Wert von 0 gibt. Es kommen also in jedem der 100 Romane des Kernkorpus Raumthemen vor.

Dass die vier Indexwerte grundsätzlich unterschiedliche Größenordnungen einnehmen, zeigt Abbildung 9.5. Während Raumthemen durchschnittlich weniger als 0,5 % eines Romans ausmachen (nämlich 0,34 %), sind rund 2 % eines Romans im

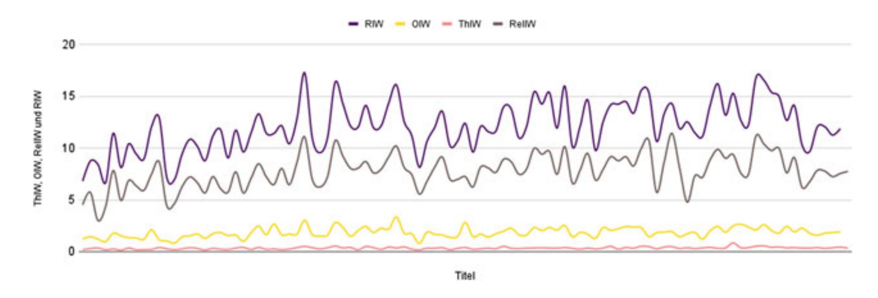


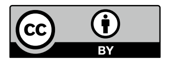
Abbildung 9.5 ThIWs, OIWs, RelIWs und RIWs im gesamten Kernkorpus

Schnitt Ortsnennungen. Die Darstellung des relationalen Raums macht rund 8 % von Romanen aus. Geht man von einem komplexen, relativ weit gefassten literarischen System der Raumdarstellungen aus, so nimmt dies in der Regel etwa 13 % eines Erzähltextes ein.

In Abschnitt 6.3 wurden die Tests des Raum-Classifiers ausführlich ausgewertet. An dieser Stelle wird nun erneut relevant, wie zuverlässig die Ergebnisse der automatischen Annotation in Bezug auf die einzelnen Korpora sind. Mit dem RIW werden alle Raumkategorien mit einbezogen, das heißt, hier gibt es eine gewisse Varianz in der Erkennungsgenauigkeit der Kategorien, die sich allerdings nicht für die einzelnen Korpora spezifizieren lässt. Das heißt, dass z. B. im 18. Jahrhundert nicht generell alle Raumkategorien schlechter erkannt werden (was zu einer Verzerrung der RIWs führen könnte, sodass eine Entwicklung angenommen werden könnte, die in Wahrheit nur eine Inkonsistenz der Erkennungsgenauigkeit ist). Mit den relationalen Kategorien, Orten und Raumthemen wurden hier ergänzend vier Kategorien ausgewählt, die eine vergleichsweise hohe durchschnittliche Erkennungsgenauigkeit zwischen rund 74 und 80 % erreichen. In den acht untersuchten Testtexten sind allerdings erhebliche Schwankungen erkennbar. Die mit Abstand größte Schwankung weist die Kategorie der Raumbeschreibungen auf, die bei Andreas-Salomé *Ruth* (19. Jahrhundert) rund 82 % F1-Score erreicht, bei Huber *Luise* (18. Jahrhundert) aber nur 42 %. Die kleinste Varianz gibt es mit einer maximalen Differenz von 14 % zwischen Schiller *Der Geisterseher* und Huber *Luise* (beides 18. Jahrhundert). Ebenso wie bei den Testtexten kann also die Genauigkeit der Erkennung auch bei den Texten des Kernkorpus pro Kategorie erheblich schwanken. Um einen Hinweis darauf zu bekommen, wie belastbar die hier beobachteten Phänomene trotz dieser Schwankungen sind, wurde eine zweite Analyse zur Cross-Validierung herangezogen. Bei dieser wurde mit einer Vorversion des Tools, das insgesamt eine durchschnittliche Erkennungsgenauigkeit von 74,77 % erreicht, also etwa ein Pro-

zent geringer als bei der Endversion, die komplette Auswertung, die in den Kapiteln 8 und 9 dokumentiert ist, noch einmal wiederholt. In Abschnitt 8.2.3 wurde bereits erwähnt, dass mit Hilfe dieser zweiten Analyse festgestellt werden konnte, dass die Annotationszahlen der weniger häufig annotierten Wörter zum Teil ausreichend stark abweichen können, um das Ranking leicht zu verändern. Bei der Berechnung der Indexwerte konnte mit Hilfe dieser zweiten Analyse beobachtet werden, dass die Annotationswerte der einzelnen Kategorien nach verwendetem Tool zwar erheblich abweichen können, drei Aspekte führen aber dazu, dass diese Form der Auswertung vergleichsweise stabil bleibt. Zunächst einmal sind – zumindest beim RIW – die Werte der Einzelkategorien nur ein Teil des betrachteten Gesamtgefüges und fallen darum nicht zu sehr ins Gewicht. Außerdem werden diese Werte normalisiert, indem sie in Relation zur Gesamtwortzahl der Romane gesetzt werden. Schließlich werden die einzelnen Werte nicht einzeln, sondern im Gesamtgefüge des Kernkorpus angeschaut. Die beobachteten Trends können also durchaus als stabil angesehen werden. Auch die Dopplung eines Textes im Kernkorpus mit Trainingsmaterial hat sich während der Analysen nicht als Störfaktor erwiesen. Der zur Beobachtung von eventuellen Verzerrungen eingefügte Text *Hyperion Fragment* von Hölderlin hat sich während der Auswertung weder bei der Betrachtung der Indexwerte noch bei der Interpretation der Annotationsdaten als auffällig hervor getan.

Ergänzend zu den beispielhaften Analyseansätzen, die vom Distant-Reading ausgehen, dann in die Korpora hereinzoomen und schließlich einzelne Kontexte genauer fokussieren, wurde hier noch ein mathematisch-statistischer Ansatz entwickelt, der das Kernkorpus ausschließlich aus der Distanz betrachtet. Dadurch treten diachrone Entwicklungen besonders gut hervor. Die Berechnung von Indexwerten hat den großen Vorteil, dass dadurch ein Eindruck gewonnen werden kann, wie wichtig die Darstellung von Raum in literarischen Texten ist. Die Indikatoren literarischer Raumdarstellung werden zu einer messbaren Größe zusammengerechnet. Gleichzeitig sagen die einzelnen Raum-Index-Werte für sich genommen nur sehr wenig aus. Erst im Vergleich mit anderen Texten, Durchschnittswerten und Standardabweichungen wird klar, ob ein Text durch besonders ausführliche Raumdarstellungen hervorsticht. Wenn die hier betrachtete Stichprobe mit 100 Texten insgesamt auch relativ klein ist, so trägt diese Studie doch dazu bei, Vergleichswerte zu verzeichnen, die bei Anschlussuntersuchungen herangezogen werden können, um Gegenstände literaturwissenschaftlicher Raumforschung abzugleichen und einordnen zu können.


BY



10 Fazit und Ausblick

Referenzierungen von Raum sind eine überzeitliche Konstante in Erzähltexten des 18.-21. Jahrhunderts. Und doch lassen die Daten, die während dieser Analyse generiert und betrachtet wurden, auch den Schluss zu, dass die Bedeutung von Raum für die Gegenwartsliteratur heute – wenn auch nur minimal – tendenziell größer ist als sie es im 18. Jahrhundert zu Lessings Lebzeiten noch war. Die dieser Arbeit zu Grunde liegende Hypothese, dass zu Lessings Zeiten Raumdarstellungen einen kleineren Anteil an Erzähltexten hatten als an zeitgenössischen und dass die Entwicklung zu stärker von Raumdarstellungen dominierten Erzähltexten diachron verlief, konnte also verifiziert werden. Doch im Prozess dieser Forschungsarbeit konnten noch drei weitere Teilergebnisse erarbeitet werden, die abschließend zusammengefasst und in ihrer Relevanz für die literaturwissenschaftliche Forschung bewertet werden.

Im ersten Teil dieser Arbeit stand die Auswertung traditionell-analoger Forschungsbeiträge im Vordergrund. Diese wurden zu einem theoriebasierten Modell zusammengefasst. Dabei handelt es sich um ein System aus sechs in ihrer Qualität sehr unterschiedlichen Kategorien, die zusammen für ein skalar-vektorielles Raumverständnis stehen. Dieses fuzzy-set-Modell kann – so hat es die Implementierung im zweiten Teil der Arbeit gezeigt – trotz der Heterogenität der Unterkategorien so operationalisiert werden, dass es zur Basis eines Machine-Learning-Trainings werden kann, das dazu führt, dass Wörter, die Raum referenzieren, automatisch vom Algorithmus erkannt und klassifiziert werden. Im Prozess des Trainings, der Tests und der Bewertung der Resultate hat sich aber auch gezeigt, dass Raumdarstellungen interpretativ und häufig sehr unterschiedlich klassifiziert werden. Dabei sind sich manche der Raumkategorien näher, während andere mehr Trennschärfe aufweisen. Erstaunlich ist der Vergleich des Inter-Annotator-Agreements mit der Erkennungsquote des Raum-Classifiers. Hier zeigt sich, dass die automatische Annotation im Vergleich mit meiner manuellen Test-Annotation eine sehr ähnliche

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5_10

Übereinstimmungsquote erreicht wie zwei voneinander unabhängige menschliche Annotatorinnen. Einen Text von einem Computer annotieren zu lassen, anstatt es selbst zu tun, führt also zu einer – rein quantitativ – ähnlichen Datenbasis wie die Annotation durch eine*n andere*n Literaturwissenschaftler*in.

Das zweite Teilergebnis neben dem theoriebasierten Raummodell umfasst gleich drei Dinge. Im Vordergrund der in dieser Arbeit angewendeten Methode stand das Machine-Learning-Training eines StanforNER-Raum-Classifiers, also einer Anwendung mit Hilfe derer Indikatoren für Raumdarstellung in Erzähltexten automatisch erkannt und entsprechend des Raummodells klassifiziert werden können. Diese Software wurde auf Zenodo publiziert und damit zur Nachnutzung bereit gestellt. Eher aus der Not heraus, Raummetaphern mit dieser Methode nicht angemessen in die Betrachtungen integrieren zu können, entstand die relationale (Meta-) Datenbank der Raummetaphern *laRa*. Hierin sind alle rund 800 während des Machine-Learning-Trainings erfassten raummetaphorischen Ausdrücke verzeichnet und mit den Erzähltexten verknüpft, in denen sie vorkommen. *LaRa* eignet sich sehr gut zur Recherche von Raummetaphern und ihrem diachronen Wandel. Obwohl die Datenbasis relativ klein ist, konnten die Analysen dieser Studie belegen, dass sie einen guten Einblick in die Nutzung von Raummetaphern in Erzähltexten der Jahrhunderte 18–21 bietet. Auch dieses (Neben-)Produkt der vorliegenden Arbeit wurde veröffentlicht und kann als Ressource frei nachgenutzt werden. Zur Berechnung von Indexwerten, die den relationalen Vergleich quantitativer Aspekte von Raumdarstellungen im diachronen Verlauf ermöglichen, wurde eine einfache Formel entwickelt. Diese kann auf das ganze Raum-Kategorien-System, auf Unterkomplexe oder einzelne Kategorien angewendet werden. Diese drei Tools und methodischen Umsetzungen der theoretischen Überlegungen dieser Studie sind in ihrer Innovation vor allem für die digitale Literaturwissenschaft relevant. Bisher gab es noch kein Verfahren, mit Hilfe dessen Raumdarstellungen in deutschsprachigen Erzähltexten automatisch erkannt und anhand heterogener Unterkategorien klassifiziert werden konnten. Die hier entwickelte Software bietet vor allem einen ersten Schritt zur massenweisen Erkennung von Wörtern, die zu Raumdarstellungen beitragen. Ist diese Masse der Raumausdrücke erkannt, so können im nächsten Schritt z. B. durch weitere Strukturierungen der Annotationsdaten tiefer gehende narratologische Analysen erstellt werden. Ähnliches gilt für *laRa*. Die digitale literaturwissenschaftliche Metaphernforschung hat bisher noch nicht dazu geführt, dass diese in Texten automatisch erkannt werden können. Auch das Machine-Learning-Training, das Teil dieser Untersuchung war, zeigt hier keine vielversprechenden Ergebnisse. Es gab aber bisher auch kein digitales Verzeichnis literarischer Raummetaphern, sodass auch hier ein erster Schritt getan werden konnte. Auf Basis der in *laRa* verzeichneten Metaphern könnte durch Erweiterung der Datenbasis z. B. durch einen Zusam-

menschluss mit anderen Projekten der digitalen Metaphernforschung ein erneuter Versuch unternommen werden, um eine kritische Masse an Raummetaphern-Daten zusammenzustellen und so einen Teil dazu beitragen, einer maschinellen Erfassung den Weg zu bereiten. Für die Formel zur Errechnung von Raumindexwerten wäre eine Implementierung (vielleicht zusammen mit dem Raum-Classifier) denkbar, die diese Werte als Metrik der räumlichen Dimension von Erzähltexten automatisch errechnet. Die in der Analyse in Kapitel 9 aufgeführten Indexwerte für Texte aus vier Jahrhunderten könnten dann als Vergleichswerte zur Kontextualisierung herangezogen werden.

Neben der zentralen Fragestellung war für diese Arbeit auch der Gedanke leitend, dass traditionelle literaturwissenschaftliche Forschung mit neuen digitalen Methoden verknüpft werden kann, um zu einer fruchtbaren Synthese zu gelangen. Das dritte und vielleicht bedeutendste Teilergebnis dieser Arbeit ist, dass diese Verknüpfung gelingen kann. Der dritte Teil dieser Studie zeigt in einer Reihe von Einzelanalysen, dass es möglich ist, mit digital unterstützten Studien direkt an traditionell-analoge Forschung der Literaturwissenschaften anzuschließen. Es können sowohl theoretische Konzepte weitergedacht als auch einzelne Thesen geprüft, gestützt und manchmal auch widerlegt werden. Auch der explorative Zugang zur hier mit Hilfe digitaler Tools und Methoden erstellten Datenbasis kann zu Einsichten in einzelne Phänomene kleinerer Teilkorpora führen. Diese intra- und intertextuelle (im weiten Wortsinne) Perspektive erlaubt es, eine größere Distanz zum einzelnen Phänomen aufzubauen und z. B. mehrere Raumthemen zugleich in den Blick zu nehmen. Unabhängig davon, ob eine enge oder eher eine weite Verknüpfung mit analog-traditioneller literaturwissenschaftlicher Forschung gewählt wird, sollten die hier angewendeten Methoden stets in dem Bewusstsein eingesetzt werden, dass sie anfällig für Verzerrungen sind. Wie Ausreißer-Phänomene zu solchen Verzerrungen führen können, wurde an Beispielen aus dem Kernkorpus gezeigt.

Mit neuen Methoden über alte Fragen nachdenken und so zu neuen Einsichten gelangen – dieses Ziel konnte in dieser Studie in mehrfacher Hinsicht erreicht werden. Nun ist es aber kein Endziel, sondern eher die Grundlage einer Reihe von neuen Studien-Möglichkeiten, die sich zum Teil schon während der Arbeit an diesem Buch ergeben haben. Raumdarstellungen mit anderen narratologisch relevanten Kategorien, wie z. B. Figuren zu verknüpfen, erscheint mir dabei eine besonders fruchtbare Möglichkeit zu sein. Das mit Fokus auf Genderzuschreibungen zu tun, die ebenfalls automatisiert annotiert werden können, ist bereits geplant. Auch die Übertragung der hier entwickelten Methode auf andere literarische Textarten wie z. B. Dramen ist vielversprechend. Vor allem, da diese Textform einerseits Lessing als Autor von Dramen noch näher ist und andererseits andere nicht oder weniger narrative Möglichkeiten der Raumdarstellung zur Verfügung hat, wäre ein Test des Raum-

Classifiers und des dahinter stehenden theoriebasierten Modells sehr interessant. In dieser Studie wurde ein (im Vergleich zu anderen Digital-Humanities-Studien) vergleichsweise kleines Korpus betrachtet. Nun, da der Raum-Classifier annehmbare Resultate erreicht, wäre es aber von großem Interesse, auch die Skalierbarkeit der Anwendung zu testen. Mehr Erzähltexte in den Blick zu nehmen und damit noch mehr Annotationsdaten zu generieren – führt das eigentlich zu neuen Einsichten oder bloß dazu, dass irgendwann der Blick für das einzelne Phänomen verloren geht? Auf diese Grundsatzfrage des Distant Reading mit immer größer werdenden Datensätzen könnte mit Hilfe der Methoden dieser Studie möglicherweise eine Antwort gefunden werden. Obwohl die zu Beginn dieses Buches aufgestellte Hypothese validiert werden konnte, führen die Teilergebnisse, Modelle, Tools, Methoden und die Synthese von Tradition und Innovation, die im Laufe dieser Studie erarbeitet wurden, keineswegs dazu, dass die Forschung zur literarischen Darstellung von Raum sich einem Abschluss nähert. Im Gegenteil eröffnen sich eine Fülle neuer Möglichkeiten, um Raumphänomene in der Literatur auch weiterhin auf neue Weise zu betrachten.

Literatur

Abbott, P. (2011). Narrativity. In Hühn, P., Meister, J.-Ch., Pier, J. und Schmid, W. The living handbook of narratology. Hamburg. https://www.lhn.uni-hamburg.de/node/27.html [24.4.2022].

Aristoteles (1995). *Physikvorlesung*. Akademie Verlag, Berlin.

Artstein, R. (2017). Inter-annotator Agreement. In Ide, N. and Pustejovsky, J., editors, *Handbook of Linguistic Annotation*, pages 297–313. Springer Netherlands, Dordrecht.

Artstein, R. and Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.

Assmann, A. (2013). Theorien des kulturellen Gedächtnisses. In Rippl, G. und Winko, S., Herausgeber, *Handbuch Kanon und Wertung: Theorien, Instanzen, Geschichte*, Seiten 76–84. Springer, Berlin.

Assmann, A. (2015). Geschichte findet Stadt. In *Geschichte findet Stadt*, Seiten 13–28. Transcript Verlag, Bielefeld.

Bachelard, G. (1987). *Poetik des Raumes*. Fischer Wissenschaft. Fischer Taschenbuch, ungek. ausg edition, Frankfurt am Main.

Bachtin, M. (2008). *Chronotopos*. Suhrkamp-Taschenbuch Wissenschaft. Suhrkamp, 1. aufl. edition, Frankfurt am Main.

Bal, M. (1985). *Narratology*. University of Toronto Press, Toronto.

Bamman, D., Popat, S., and Shen, S. (2019). An annotated dataset of literary entities. In *Proceedings of the 2019 Conference of the North*, pages 2138–2144, Minneapolis, Minnesota. Association for Computational Linguistics.

Baradaran, R., Ghiasi, R., and Amirkhani, H. (2020). A Survey on Machine Reading Comprehension Systems. *arXiv:2001.01582 [cs]*. arXiv: 2001.01582.

Barth, F. und Viehhauser, G. (2017). Digitale Modellierung literarischen Raums. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4622733.

Bastian, M., Heymann, S., and Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks.

Bauer, M. (1997). *Romantheorie*. Bd. 305 in Sammlung Metzler. J.B. Metzler, Stuttgart.

Becker, O. und Patzig, G. (1966). *Das mathematische Denken der Antike.* Vandenhoeck & Ruprecht, Göttingen.

Bieker, N. (2019). *Erzählanfänge und Erzählschlüsse im Adoleszenzroman*. Band 118 in Kinder- und Jugendkultur, -literatur und -medien. Peter Lang, Berlin.

M. K. Schumacher, *Orte und Räume im Roman*, Digitale Literaturwissenschaft,
https://doi.org/10.1007/978-3-662-66035-5

Blumenberg, H. (1971). Beobachtungen an Metaphern. *Archiv für Begriffsgeschichte*, 15:161–214. Publisher: Temporary Publisher.

Blumenberg, H. (2001). *Ästhetische und metaphorologische Schriften*. Band 1513 in Suhrkamp Taschenbuch Wissenschaft. Suhrkamp, Frankfurt am Main, 1. aufl.

Blumenberg, H. (2013). *Paradigmen zu einer Metaphorologie*. Band 10 in Suhrkamp Studienbibliothek. Suhrkamp, Frankfurt am Main, 1. aufl.

Bodenhamer, D. J., Corrigan, J., and Harris, T. M., editors, (2010). *The spatial humanities: GIS and the future of humanities scholarship*. Spatial humanities. Indiana University Press, Bloomington.

Borin, L., Kokkinakis, D., and Olsson, L.-J. (2007). Naming the Past: Named Entity and Animacy Recognition in 19th Century Swedish Literature. In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007).*, pages 1–8, Prague, Czech Republic. Association for Computational Linguistics.

Bourdieu, P. (2006). Sozialer Raum, symbolischer Raum. In Dünne, J. und Günzel, St., Herausgeber, *Raumtheorie. Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seiten 354–366. Suhrkamp, Frankfurt am Main.

Brooke, J., Hammond, A., and Baldwin, T. (2016). Bootstrapped Text-level Named Entity Recognition for Literature. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers).

Burk, S. (2018a). Einleitung. In Burk, St., Klepikova, T. und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*. Peter Lang, Berlin; New York.

Burk, S. (2018b). Privatheit und Öffentlichkeit als semantische Oppositionen. Vorbemerkungen. In Burk, St., Klepikova, T. und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*. Peter Lang, Berlin; New York.

Butor, M. (1968). *Probleme des Romans*. Beck'sche schwarze Reihe. Beck, München.

Bögel, T., Gertz, M., Gius, E., Jacke, J., Meister, J. C., Petris, M., and Strötgen, J. (2015). Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.3240591

Cassirer, E. (2006). Mythischer, ästhetischer und theoretischer Raum. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 485. Suhrkamp, Frankfurt am Main.

Certeau, M. d. (2006). Praktiken im Raum. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 343. Suhrkamp, Frankfurt am Main.

Chatman, S. (1993). *Story and Discourse Narrative Structure in Fiction and Film*. Publisher: Cornell University Press, New York.

Clarke, S. (1717a). Dr. Clarke's Fifth Reply (Normalized Version). https://web.archive.org/web/20101211174753/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00235 [30.11.2022].

Clarke, S. (1717b). Dr. Clarke's Third Reply (Normalized Version). https://web.archive.org/web/20120524180657/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00231 [30.11.2022].

Cohen, M. (2015). Atlantik/Pazifik die imaginäre Erschließung der Ozeane im Zeitalter der Segelschifffahrt. In Dünne, J. und Mahler, A., Herausgeber, Handbuch Literatur & Raum.

Seiten 364–375. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Csáky, M. und Leitgeb, C. (2009). Kommunikation – Gedächtnis – Raum: Orientierungen im spatial turn der Kulturwissenschaften. In Csáky, M. und Leitgeb, C., Herausgeber, *Kultur- und Medientheorie*, Seiten 7–10. transcript Verlag, Bielefeld, Germany, 1 aufl.

Czezior, P. (2018). Private Innerlichkeit als Gegenentwurf zur Gefahr des Gefangenseins im öffentlichen Alltag des Philisters. E.T.A. Hoffmanns „Goldener Topf" und die Metamorphose des Anselmus. In Burk, St., Klepikova, T. und Piegsa, M., Herausgeber, *Privates Erzählen. Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*. Seiten 195–218. Peter Lang, Berlin.

Deleuze, G. (2006). Das Glatte und das Gekerbte. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 434. Suhrkamp, Frankfurt am Main.

Dennerlein, K. (2009). *Narratologie des Raumes*. Narratologia 22. De Gruyter, Berlin.

Descartes, R. (2007). *Die Prinzipien der Philosophie*, Ausgabe 566. Felix Meiner Verlag, Leipzig.

Didakowski, J., Geyken, A., und Hanneforth, T. (2007). Eigennamenerkennung zwischen morphologischer Analyse und Part-of-Speech Tagging: ein automatentheoriebasierter Ansatz. *Zeitschrift für Sprachwissenschaft*, 26(2).

Dobson, J. E. (2019). *Critical Digital Humanities: The Search for a Methodology*. Topics in the digital humanities. University of Illinois Press, Urbana, Illinois.

Doering, S. (2013). Paradies und Idylle. In Breuer, I., Herausgeber, *Kleist- Handbuch. Leben – Werk – Wirkung*, Seiten 351–354. Metzler, Stuttgart.

Doležel, L. (2009). *Heterocosmica: fiction and possible worlds*. Parallax revisions of culture and society. Johns Hopkins University Press, Baltimore.

Dröscher-Teille, M. (2018). Privatheit als „male oscuro". Ingeborg Bachmanns Malina und die Aufzeichnungen aus der Zeit der Krankheit zwischen Diskretion, Voyeurismus und Tabubruch. In Burk, St,. Klepikova, T. und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*, Seiten 297–314. Peter Lang, Berlin.

Dziudzia, C. und Hall, M. (2020). Die Kanonfrage 2.0. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4621782.

Döring, J. (2008). Einleitung Was lesen wir im Raume? Der Spatial Turn und das geheime Wissen der Geographen. In Dörig, J. und Thielmann, T., Herausgeber, Spatial Turn : das Raumparadigma in den Kultur- und Sozialwissenschaften, Seiten 7–48. Transcript Verlag, Bielefeld.

Dünne, J. (2015). Dynamisierungen Bewegung und Situationsbildung. In Dünne, J. and Mahler, A., Herausgeber, *Handbuch Literatur & Raum*, Seiten 41–56, Handbücher zur kulturwissenschaftlichen Philologie 3. DeGruyter, Berlin; Boston.

Dünne, J. und Mahler, A., Herausgeber (2015). *Handbuch Literatur & Raum*. Band 3 in Handbücher zur kulturwissenschaftlichen Philologie. De Gruyter, Berlin; Boston.

Ehrlicher, H. (2015). Die „Neue Welt" Reisen und Alterität. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 355–363. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Einstein, A. (2006). Raum, Äther und Feld in der Physik. In Dünne, J. und Günzel, S., Herausgeber, Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften, Seite 94. Suhrkamp, Frankfurt am Main.

Erlebach, P. (1990). *Theorie und Praxis des Romaneingangs Untersuchungen zur Poetik des Englischen Romans*. Anglistische Forschungen 208. Winter, Heidelberg.

Euclides (1971). *Die Elemente Buch I–XIII*. Wissenschaftliche Buchgesellschaft, 4., unveränd. aufl. reprograf. nachdr. [d. ausg. leipzig, akad. verl.- ges., 1933–1937], Darmstadt.

Faruqui, M. and Padó, S. (2010). Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In *KONVENS*.

Filler, A. (1993). *Euklidische und nichteuklidische Geometrie*. Nummer 7 in Mathematische Texte. BI-Wiss.-Verl, Mannheim.

Finkel, J. R., Grenager, T., and Manning, C., editors, (2005). Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics*, pages 363–370.

Fludernik, M. (2011). Conversational Narration – Oral Narration. In Hühn, P., Meister, J.-Ch., Pier, J. und Schmid, W., editors, *The living handbook of narratology*. Hamburg. https://www.lhn.uni-hamburg.de/node/34.html [24.4.2022].

Flusser, V. (2006). Räume. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seiten 274–288. Suhrkamp, Frankfurt am Main.

Foucault, M. (2005). *Die Heterotopien zwei Radiovorträge; [7. und 21. Dezember 1966] = Les hétérotopies*. Suhrkamp, zweisprachige ausg., 1. aufl. Frankfurt am Main.

Foucault, M. (2006). Von anderen Räumen. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 317. Suhrkamp, Frankfurt am Main.

Foucault, M. (2008). Die Ordnung der Dinge. In *Die Hauptwerke*, Seiten 7–470. Suhrkamp, Frankfurt am Main.

Frank, J. (1945a). Spatial Form in Modern Literature, Part I. *The Sewanee review.*, 53:221.

Frank, J. (1945b). Spatial Form in Modern Literature, Part II. *The Sewanee review.*, 53:433.

Frank, J. (1977). Spatial Form: An Answer to Critics. *Critical Inquiry*, 4(2):231–252. Publisher: The University of Chicago Press.

Freise, M., Starre, A., Beilein, M., und Assmann, A. (2013). Literaturwissenschaftliche Kanontheorien und Modelle der Kanonbildung. In Rippl, G. und Winko, S., Herausgeber, *Handbuch Kanon und Wertung: Theorien, Instanzen, Geschichte*, Seiten 50–84. J.B. Metzler, Stuttgart.

Freudenstein, C., Herausgeber (2016). *Deutsche Literatur der Gegenwart*. Kindler Kompakt. J.B. Metzler Verlag, Stuttgart.

Genette, G. (1976). *Figure*. Piccola biblioteca Einaudi. Einaudi, Turin.

Genette, G. (1998). *Die Erzählung*, volume 8083 of *UTB Literatur- und Sprachwissenschaft*. Fink, München, 2. aufl.

Gerstorfer, D. (2020). Entdecken und Rechtfertigen in den Digital Humanities. In Reiter, N., Pichler, A., und Kuhn, J., Herausgeber, *Reflektierte algorithmische Textanalyse*, Seiten 107–124. De Gruyter, Berlin.

Giacobazzi, C. (2016). Die Dialektik von Himmel und Erde zwischen Klassik, Romantik und Realismus am Beispiel von den Wahlverwandtschaften, Heinrich von Ofterdingen und Immensee. In Lesch, H., Oberdorfer, B., und Waldow, S., Herausgeber, *Der Himmel als*

transkultureller ethischer Raum: Himmelskonstellationen im Spannungsfeld von Literatur und Wissen. Vandenhoeck & Ruprecht, Göttingen.

Gius, E. und Jacke, J. (2015). Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. Publisher: HAB – Herzog August Bibliothek, Wolfenbüttel. DOI https://doi.org/10.17175/SB001_006.

Gius, E. and Jacke, J. (2017). The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis. *International Journal of Humanities and Arts Computing*, 11(2):233–254.

Gius, E., Meister, J. C., Meister, M., Petris, M., Bruck, C., Jacke, J., Schumacher, M., Gerstorfer, D., Flüh, M., and Horstmann, J. (2022). CATMA. DOI https://doi.org/10.5281/ZENODO.1470118.

Goeth, S. M. T. (2018). Private Räume und intimes Erzählen bei Novalis. In Burk, St., Klepikova, T. und Piegsa, M., Herausgeber, *Privates Erzählen*. Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts. Seiten 47–70. Peter Lang, Berlin.

Gregory, I. N. and Geddes, A. (2014). *Toward spatial humanities: historical GIS and spatial history. The spatial humanities*. Indiana University Press, Bloomington.

Gromes, H. (2005). Bertolt Brechts Chinoiserien: Von der ‚Maßnahme' bis zu ‚Der gute Mensch von Sezuan'. In Wintgens, H.-H., Herausgeber, *Literarische Orte – Orte Der Literatur*, Seiten 151–167. Universitätsverlag, Hildesheim.

Gödde, S. (2015). Athen Autochthoner Raum und politisches Theater. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 312–323. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Günzel, S. (2008). Spatial Turn – Topographical Turn – Topological Turn über die Unterschiede zwischen Raumparadigmen. In Dörig, J. und Thielmann, T., Herausgeber, *Spatial Turn: das Raumparadigma in den Kultur- und Sozialwissenschaften*, Seiten 219–240. Transcript Verlag, Bielefeld.

Günzel, S. (2009a). Einleitung. In Günzel, S., Herausgeber, *Raumwissenschaften*, Seiten 7–13. Suhrkamp, Frankfurt am Main.

Günzel, S., Herausgeber (2009b). *Raumwissenschaften*. Nummer 1891 in Suhrkamp Taschenbüch Wissenschaft. Suhrkamp, Frankfurt am Main, 1. aufl.

Hafner, M. (2020). Der ‚mythos athen' im literarischen diskursfeld fiktionaler erzählprosa der kaiserzeit. lukian, chariton und heliodor. In Cecconi, P. und Tornau, C., Herausgeber, *Städte und Stadtstaaten zwischen Mythos, Literatur und Propaganda*, Seiten 249–268. De Gruyter, Berlin.

Hammann, M., Jördens, J., und Schecker, H. (2014). Übereinstimmung zwischen Beurteilern: Cohens Kappa, https://docplayer.org/32363879-Uebereinstimmung-zwischen-beurteilern-cohens-kappa-k.html [29.4.2022].

Hegel, G. W. F. (1979). Enzyklopädie der philosophischen Wissenschaften im Grundrisse, Zweiter Teil: Die Naturphilosophie, 1. Abteilung: Die Mechanik, A. Raum und Zeit, a. Der Raum, http://www.zeno.org/Philosophie/M/Hegel,+Georg+Wilhelm+Friedrich/Enzyklop%C3%A4die+der+philosophischen+Wissenschaften+im+Grundrisse/Zweiter+Teil%3A+Die+Naturphilosophie/1.+Abteilung%3A+Die+Mechanik/A.+Raum+und+Zeit/a.+Der+Raum [29.4.2022].

Heidegger, M. (2006). Die Räumlichkeit des Daseins. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 141. Suhrkamp, Frankfurt am Main.

Herman, D. (2012). Exploring the Nexus of Narrative and Mind. *Narrative theory*, page 14.
Herman, D., editor (2007). *The Cambridge companion to narrative*. Cambridge companions to literature. Cambridge University Press, Cambridge; New York, 1. publ. edition.
Herman, D., editor (2010). *Routledge encyclopedia of narrative theory*. Routledge, London, repr edition.
Heßbrüggen-Walter, S. (2017). Modellierung: eine Begriffsbestimmung. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4645271.
Hoffmann, G. (1978). *Raum, Situation, erzählte Wirklichkeit poetologische und historische Studien zum englischen und amerikanischen Roman*. Metzler [u. a.], Stuttgart, 1. aufl.
Horstmann, J. (2018). Topic Modeling. In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT - Literatur digital erforschen*. https://fortext.net/routinen/methoden/topic-modeling [29.4.2022].
Horstmann, J. (2019). DROC: Deutsches Romankorpus. In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT - Literatur digital erforschen*. https://fortext.net/ressourcen/textsammlungen/droc-deutsches-romankorpus [29.4.2022].
Horstmann, J. und Kern, A. (2018). Deutsches Textarchiv (DTA). In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT - Literatur digital erforschen*. https://fortext.net/ressourcen/textsammlungen/deutsches-textarchiv-dta [29.4.2022].
Huber, M. (2017). Die Erinnerungs-Kartografie der Provinz samt ‚Es war einmal...'. Saša Stanišićs Roman Vor dem Fest als Versuch einer intersubjektiven Gegenwartsfindung. In Iztueta, G., Saalbach, M., Talavera, I., Bescansa, C. und Standke, J., Herausgeber, *Raum–Gefühl–Heimat: Literarische Repräsentationen nach 1945*, Seiten 157–175. LiteraturWissenschaft.de, Marburg.
Huber, S. (2018). „Aber privat sein war so gar nicht sein Fall". Räume des Privaten in den Überwachungsromanen Corpus Delicti von Juli Zeh und Fremdes Land von Thomas Sautner. In Burk, S., Klepikova, T., und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*, Seiten 195–218. Peter Lang, Berlin; New York.
Husserl, E. (1991). *Ding und Raum Vorlesungen 1907*. Philosophische Bibliothek 437. Meiner, Hamburg.
Hänig, C., Bordag, S., and Thomas, S. (2014). Modular Classifier Ensemble Architecture for Named Entity Recognition on Low Resource Systems. Seiten 113–116, Hamburg, https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxnZXJtZXZhbDIwMTRuZXJ8Z3g6NTk4M2UzZWMzYjM1NTY0Zg [19.2.2020].
Jacke, J. (2018). Manuelle Annotation. In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT - Literatur digital erforschen*. https://fortext.net/routinen/methoden/manuelle-annotation [29.4.2022].
Jackson, J. E. (2003). *A user's guide to principal components*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J.
Janka, M. (2015). Das Mittelmeer Reisen, Navigieren, Erzählen. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 301–311. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.
Jannidis, F. (2017). Grundlagen der Datenmodellierung. In Jannidis, F., Kohle, H., und Rehbein, M., Herausgeber, *Digital Humanities*, Seiten 99–108. J.B. Metzler, Stuttgart.

Jannidis, F., Krug, M., Reger, I., Toepfer, M., Weimer, L., und Puppe, F. (2015). Automatische Erkennung von Figuren in deutschsprachigen Romanen. In *Von Daten zu Erkentnissen*, Seiten 1–6, Graz.

Jiang, R., Banchs, R. E., and Li, H. (2016). Evaluating and Combining Name Entity Recognition Systems. In *Proceedings of the Sixth Named Entity Workshop*, pages 21–27, Berlin, Germany. Association for Computational Linguistics.

Jing, X. (2015). Vom Artushof nach King's Landing Chrestien de Troyes und die Mediävisierung politischer Topographie. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 335–343. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Jürjens, K. (2018). Flecken und Falten. Novellistische Spurensuche im Interieur (Gottfried Kellers Regine). In Kunz, E. A. und Légeret, J., Herausgeber, *Raum und Narration*, Seiten 29–44. Aisthesis Verlag, Bielefeld.

Kant, I. (1993). *Kritik der reinen Vernunft*. Kant, Immanuel 1724–1804 Die drei Kritiken. Felix Meiner Verlag, Hamburg.

Kant, I. (2006). Von dem ersten Grunde des Unterschiedes der Gegenden im Raum. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 74. Suhrkamp, Frankfurt am Main.

Klotz, V. (1969). *Die erzählte Stadt ein Sujet als Herausforderung des Romans von Lesage bis Döblin*. Hanser, München.

Knode, F. (2018). Privatheit(en) unter dem Vorzeichen der Empfindsamkeit. Modellierungen von Ich-Entfaltungen und Ich-Isolation in Friedrich Heinrich Jacobis Woldemar. In Burk, S., Klepikova, T., und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*. Seiten 23–46. Peter Lang, Berlin; New York.

Kopernikus, N. (2007). *De revolutionibus die erste deutsche Übersetzung in der Grazer Handschrift: kritische Edition*. Kopernikus, Nikolaus 1473–1543 Gesamtausgabe. Akademie Verlag, Berlin.

Krobb, F. U. (2009). *Erkundungen im Überseeischen Wilhelm Raabe und die Füllung der Welt*. Königshausen u. Neumann, Würzburg.

Kuhn, J. (2018). Computerlinguistische Textanalyse in der Literaturwissenschaft? Oder: ‚The Importance of Being Earnest' bei quantitativen Untersuchungen. In Albrecht, A., Bernhart, T., Richter, S. und Willand, M. (Herausgeber) *Quantitative Ansätze in den Literatur- und Geisteswissenschaften: systematische und historische Perspektiven*. Seiten 11–44. de Gruyter, Berlin.

Kuhn, J. (2020). Einleitung. In Reiter, N., Pichler, A. und Kuhn, J., Herausgeber, *Reflektierte algorithmische Textanalyse*, Seiten 9–40. De Gruyter.

Kurzenberger, H. (2005). Murnau, München, Wien – Ödön von Horväths dramatische und biografische Orte. In Wintgens, H.-H., Herausgeber, *Literarische Orte – Orte Der Literatur*. Seiten 202–218. Universitätsverlag, Hildesheim.

Lahn, S. und Meister, J. C. (2008). *Einführung in die Erzähltextanalyse*. BA: Studium. Metzler, Stuttgart.

Lakoff, G. (1998). *Leben in Metaphern Konstruktion und Gebrauch von Sprachbildern*. Carl-Auer-Systeme, Verl. und Verl.-Buchh., Heidelberg, 1. aufl.

Lang, F. (2009). Archäologie. In Günzel, S., Herausgeber, *Raumwissenschaften*, Seiten 30–45. Suhrkamp, Frankfurt am Main.

Lechtermann, C. (2015). Der lange Raum die grosse kirch in der Vnderweisung der kunst des Messens der Offizin Hieronymus Rodler. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 344–354. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Lefebvre, H. (2006). Die Produktion des Raums. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 330. Suhrkamp, Frankfurt am Main.

Lehnert, G. (2014). Räume und Affekte. Boudoir und bürgerlicher Innenraum. In Lehnert, G. und Wehinger, B., Herausgeber, *Räume und Lebensstile im 18. Jahrhundert. Kunst-, Literatur-, Kulturgeschichte*, Seiten 13–32. Werhahn, Hannover.

Leibniz, G. W. (1717a). Mr. Leibnitz's Fifth Paper (Normalized Version). https://web.archive.org/web/20120524180620/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00234 [30.11.2022].

Leibniz, G. W. (1717b). Mr. Leibnitz's Fourth Paper (Normalized Version). https://web.archive.org/web/20120524180739/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00232 [30.11.2022].

Leibniz, G. W. (1717c). Mr. Leibnitz's Second Paper (Normalized Version). https://web.archive.org/web/20120528124759/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00228 [30.11.2022].

Leibniz, G. W. (1717d). Mr. Leibnitz's Third Paper (Normalized Version). https://web.archive.org/web/20120528124510/, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00230 [30.11.2022].

Leibniz, G. W. (2002). *Monadologie und andere metaphysische Schriften „Discours de métaphysique, La monadologie, Principes de la nature et de la grâce fondés en raison.“*. F. Meiner, Hamburg.

Lessing, G. E. (2012). Ästhetische Schriften. Laokoon. https://textgridrep.org/browse/rkbr_0 [29.4.2022].

Leufkens, K. E. (2020). Data Visualization using Principal Component Analysis within the Digital Humanities. https://studenttheses.uu.nl/handle/20.500.12932/35459 [29.4.2022].

Li, J., Chenliang, L., Han, R., and Aixin, S. (2020). A Survey on Deep Learning for Named Entity Recognition. *EEE Transactions on Knowledge and Data Engineering*, 1–1:1–20.

Lipinski, B. (2019). Wie die Dinge Geschichten erzählen: Das Exponat, die Biographie und die Literatur im neuen Buddenbrookhaus. *Dinge im Werk Thomas Manns*, Seiten 9–29. Publisher: Brill Fink Section: Dinge im Werk Thomas Manns. DOI https://doi.org/10.30965/9783846764268_003.

Liu, S., Zhang, X., Zhang, S., Wang, H., and Zhang, W. (2019). Neural Machine Reading Comprehension: Methods and Trends. *Applied Sciences*, 9(18):3698.

Lotman, J. M. (1973). *Die Struktur des künstlerischen Textes*. Edition Suhrkamp. Suhrkamp Verlag, Frankfurt am Main, 1. aufl.

Löw, M. und Sturm, G. (2016). Raumsoziologie. In Kessl, F. und Reutlinger, C., Herausgeber, *Handbuch Sozialraum*, Seiten 1–19. Springer Fachmedien, Wiesbaden.

Mahler, A. (2015). Topologie Primat der Relation. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 17–29. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Mani, I. (2013). *Computational modeling of narrative*. Publisher: Morgan & Claypool.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., and Gómez-Berbís, J. M. (2013). Named Entity Recognition: Fallacies, challenges and opportunities. *Computer Standards & Interfaces*, 35(5):482–489.

Martin-Rodilla, P., Pereira-Fariña, M., and Gonzalez-Perez, C. (2019). Qualifying and Quantifying Uncertainty in Digital Humanities: A Fuzzy-Logic Approach. In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 788–794, León Spain. ACM.

Martínez, M. und Scheffel, M. (2007). *Einführung in die Erzähltheorie*. C.-H.-Beck-Studium. Beck, München, 7. aufl.

Matzat, W. (2014). *Perspektiven des Romans: Raum, Zeit, Gesellschaft ein romanistischer Beitrag zur Gattungstheorie*. Verlag J.B. Metzler, Stuttgart.

Meister, J. C. (2008). *Computing Action A Narratological Approach*. Publisher: Walter de Gruyter, Berlin.

Meister, J. C. (2017). Erzählen: Eine anthropologische Universalie? In *Grundthemen der Literaturwissenschaft: Erzählen*, Seiten 88–112. De Gruyter, Berlin.

Meister, J.-C., Petris, M., Gius, E., Jacke, J., Horstmann, J., and Bruck, C. (2018). CATMA. https://zenodo.org/record/1470119 [29.4.2022].

Menzel, W. (2005). ‚Der Bach' in Wilhelm Müllers und Franz Schuberts ‚Schöne Müllerin'. In Wintgens, H.-H., Herausgeber, *Literarische Orte – Orte der Literatur*, Seiten 135–250. Universitätsverlag, Hildesheim.

Mieth, D. (2016). Der Himmel in mir. Die Interiorisierung des Himmels bei Meister Eckhart. „Was oben war, ist innen.". In Lesch, H., Oberdorfer, B., und Waldow, S., Herausgeber, *Der Himmel als transkultureller ethischer Raum*, Seiten 105–126. V&R unipress, Göttingen, 1. aufl.

Mikheev, A., Moens, M., and Grover, C. (1999). Named Entity Recognition without Gazetteers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, Norway. Association for Computational Linguistics.

Miller, N., Herausgeber (1965). *Romananfänge Versuch zu einer Poetik des Romans*. Verl. Literarisches Colloquium, Berlin.

Mitchell, W. J. T. (1980). Spatial Form in Literature: Toward a General Theory. *Critical Inquiry*, 6(3):539–567. Publisher: The University of Chicago Press.

Moretti, F. (2013). *Distant reading*. Verso, London.

Moretti, F. (2014). „Operationalizing": or, the Function of Measurement in Modern Literary Theory. *The Journal of English Language and Literature*, 60(1):3–19.

Nantke, J. (2017). *Ordnungsmuster im Werk von Kurt Schwitters zwischen Transgression und Regelhaftigkeit*. Spectrum Literaturwissenschaft Band 59. De Gruyter, Berlin.

Neo4J Graphdatabase. Graph Modeling Guidelines – Developer Guides, https://neo4j.com/developer/guide-data-modeling/ [29.4.2022].

Nitsch, W. (2015). Topographien zur Ausgestaltung literarischer Räume. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 30–40. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Nünning, A., Herausgeber (2001). *Metzler Lexikon Literatur- und Kulturtheorie: Ansätze, Personen, Grundbegriffe*. J.B. Metzler, Stuttgart, 2., überarb. und erw. aufl.

Nünning, V., Nünning, A., und Stritzke, N., Herausgeber (2004). *Erzähltextanalyse und Gender Studies*. Bd. 344 in Sammlung Metzler. Metzler, Stuttgart.

Oberdorfer, B. und Waldow, S. (2016). Einleitung. In Lesch, H., Oberdorfer, B., und Waldow, S., Herausgeber, *Der Himmel als transkultureller ethischer Raum*, Seiten 9–16. V&R unipress, Göttingen, 1. aufl.

Passavant, N. (2018). Vom Bett aus in die Moderne. Die „Urhütte" als Reflexionsraum literarischer Exzentrik. In Burk, S., Klepikova, T., und Piegsa, M., Herausgeber, *Privates Erzählen: Formen und „Funktionen von Privatheit in der Literatur des 18. bis 21. Jahrhunderts*, Seiten 129–148. Peter Lang, Berlin; New York.

Pavel, T. G. (1986). *Fictional worlds*. Harvard Univ. Press, Cambridge.

Phelan, J. (2012). Narrative Worlds: Space, Setting, Perspective. In *Narrative theory: core concepts and critical debates*, pages 3–8. Ohio State University Press, Columbus.

Piatti, B. (2008). *Die Geographie der Literatur Schauplätze, Handlungsräume, Raumphantasien*. Wallstein, Göttingen.

Pichler, A. und Reiter, N. (2020). Reflektierte Textanalyse. In Reiter, N., Pichler, A. und Kuhn, J., Herausgeber, *Reflektierte algorithmische Textanalyse*, Seiten 43–60. De Gruyter, Berlin.

Piotrowski, M. (2019). Accepting and Modeling Uncertainty. In Kuczera, A., Wübbena, Th. und Kollatz, Th., Herausgeber, *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. [Ausgewählte Beiträge der Tagung 19.-20.01.2018 an der Akademie der Wissenschaften und der Literatur*, Mainz] Zeitschrift für digitale Geisteswissenschaften / Sonderbände:4. Herzog August Bibliothek Version Number: 1.0.

Piper, A. (2018). *Enumerations data and literary study*. The University of Chicago Press, Chicago.

Piwitt, H. P. (1965). Poetische Fiktion, Wirklichkeitsauffassung und Erzählerrolle im neueren Romananfang. In Miller, N., Herausgeber, *Romananfänge. Versuch zu einer Poetik des Romans*. Literarisches Kolloquium, Berlin.

Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv:2010.16061 [cs, stat]*. arXiv: 2010.16061.

Presner, T. S. (2014). *HyperCities thick mapping in the digital humanities*. metaLABprojects. Harvard University Press, Chicago.

Reidemeister, K. (1940). *Die Arithmetik der Griechen*. Hamburger mathematische Einzelschriften 26. Teubner, Leipzig.

Reiter, N. (2020). Anleitung zur Erstellung von Annotationsrichtlinien. In Reiter, N., Pichler, A. und Kuhn, J., Herausgeber, *Reflektierte algorithmische Textanalyse*, Seiten 193–202. De Gruyter, Berlin.

Retsch, A. (2000). *Paratext und Textanfang*. Würzburger Beiträge zur deutschen Philologie 18. Königshausen & Neumann, Würzburg.

Richardson, B. (2008). Introduction: Narrative Beginnings. In Richardson, B., editor, *Narrative Beginnings: Theories and Practices*, pages 1–10. University of Nebraska Press, Lincoln.

Rippl, G. und Winko, S., Herausgeber (2013). *Handbuch Kanon und Wertung: Theorien, Instanzen, Geschichte*. Metzler, Stuttgart.

Risam, R. (2019). *New digital worlds postcolonial digital humanities in theory, praxis, and pedagogy*. Northwestern University Press, Evanston.

Rocchi, S. (2020). Triptolemos und Europa in Tarraco. Stadtdarstellung, Politik und mythische Züge in Florus' Vergilius: orator an poeta? In Cecconi, P. und Tornau, Ch., Herausgeber,

Städte und Stadtstaaten zwischen Mythos, Literatur und Propaganda, Seiten 223–248. De Gruyter, Berlin.

Rohmer, E. (2006). Film, eine Kunst der Raumorganisation. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 515. Suhrkamp, Frankfurt am Main.

Ronen, R. (1994). *Possible worlds in literary theory*. Literature, culture, theory 7. Cambridge Univ. Press, Cambridge, 1. publ. edition.

Rousseau, J.-J. (2017). Schriften zur Kulturkritik. Publisher: Felix Meiner Verlag, Hamburg.

Ryan, M.-L. (1980). Fiction, non-factuals, and the principle of minimal departure. *Poetics.*, 9(4):403.

Ryan, M.-L. (2007). Toward a definition of narrative. In Richardson, B., editor, *The Cambridge Companion to Narrative*, pages 22–38. The Cambridge University Press, Cambridge.

Ryan, M.-L. (2012a). Narration in Various Media. In Hühn, P., Meister, J.-Ch., Pier, J. and Schmid, W., editors, *The living handbook of narratology*. Hamburg. https://www.lhn.uni-hamburg.de/node/53.html [2.5.2022].

Ryan, M.-L. (2012b). Possible Worlds. In Hühn, P., Meister, J.-Ch., Pier, J. and Schmid, W., editors, *The living handbook of narratology*. Hamburg. https://www.lhn.uni-hamburg.de/node/54.html [2.5.2022].

Ryan, M.-L., Foote, K. E., and Azaryahu, M. (2016). *Narrating space/spatializing narrative: where narrative theory and geography meet*. Theory and interpretation of narrative. The Ohio State University Press, Columbus.

Römer, H. (2018). *Die norddeutsche Region in der Erzählprosa der deutschsprachigen Gegenwartsliteratur*. Philologie und Kulturgeschichte Band 8. Aisthesis Verlag, Bielefeld.

Rössler, B. (2001). *Der Wert des Privaten*. Suhrkamp-Taschenbuch Wissenschaft. Suhrkamp, Frankfurt am Main, originalausg., 1. aufl.

Sandberg, A. (2018). Ethische Ökokritik der Gegenwartsliteratur: Saša Stanišić Vor dem Fest (2014). In Dürbeck, G., Kanz, C., und Zschachlitz, R., Herausgeber, *Ökologischer Wandel in Der Deutschsprachigen Literatur Des 20. und 21. Jahrhunderts*. Peter Lang, Hamburg.

Sasse, S. (2009). Literaturwissenschaft. In Günzel, S., Herausgeber, *Raumwissenschaften*, Seiten 225–241. Suhrkamp, Frankfurt am Main.

Schmid, W. (2014). *Elemente der Narratologie*. De-Gruyter-Studium. de Gruyter, Berlin, 3., erweiterte und überarbeitete Auflage.

Schmid, W. (2008). Narrativity and Eventfulness. In *Narrativity and Eventfulness*, pages 17–34. De Gruyter, Berlin.

Schmid, W. (2017). Mentale ereignisse. In *Mentale Ereignisse*, Seiten I–IV. De Gruyter, Berlin.

Schmidt, J. C. (2009). Physik. In Günzel, S., Herausgeber, *Raumwissenschaften*, Seiten 290–307. Suhrkamp, Frankfurt am Main.

Schneider, W. (2005). „Als wär's ein Stück von mir" Carl Zuckmayer in Mainz und am Rhein. In Wintgens, H.-H., Herausgeber, *Literarische Orte – Orte der Literatur*, Seiten 219–234. Universitätsverlag, Hildesheim.

Schumacher, M. (2018a). Named Entity Recognition (NER). In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT – Literatur digital erforschen*. https://fortext.net/routinen/methoden/named-entity-recognition-ner [3.5.2022].

Schumacher, M. (2018b). Netzwerkanalyse. In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT – Literatur digital erforschen*. https://fortext.net/routinen/methoden/netzwerkanalyse [3.5.2022].

Schumacher, M. (2020a). Machine Learning: Was man beim Erstellen von Trainingsdaten (für die Literaturwissenschaft) beachten sollte. https://lebelieberliterarisch.de/machine-learning-was-man-beim-erstellen-von-trainingsdaten-fuer-die-literaturwissenschaft-beachten-sollte/ [3.5.2022].

Schumacher, M. (2020b). Netzwerkanalyse mit Gephi. In Gius, E., Schumacher, M., Gerstorfer, D. und Meister, M., Herausgeber, *forTEXT – Literatur digital erforschen*. https://fortext.net/routinen/lerneinheiten/netzwerkanalyse-mit-gephi [3.5.2022].

Schumacher, M. (2022). „Wie Wölkchen im Morgenlicht" – Zur automatisierten Metaphern-Erkennung und der Datenbank literarischer Raummetaphern laRa. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.6328183.

Schumacher, M. und Flüh, M. (2020). m*w Figurengender zwischen Stereotypisierung und literarischen und theoretischen Spielräumen. Genderstereotype und -bewertungen in der Literatur des 19. Jahrhunderts. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4621892.

Schumacher, M. K. (2021a). Forschungsdaten Orte und Räume im Roman. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4946935.

Schumacher, M. K. (2021b). laRa: Datenbank literarischer Raummetaphern. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4913253.

Schumacher, M. K. (2021c). Raum-Classifier (kompatibel mit StanfordNER). Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4922391.

Schöch, C. und Steffen, P. (2014). Für eine computergestützte literarische Gattungsstilistik. Publisher: Zenodo. DOI https://doi.org/10.5281/ZENODO.4623621.

Seeburg, U. (2018). Man soll sehen, dass hier gelebt wird. Die Inszenierung von Privatheit in literarischen Interieurs des 19. Jahrhunderts und der zeitgenössischen Interior-Fotografie. In Kunz, E. A. und Légeret, J., Herausgeber, *Raum und Narration*, Seiten 45–61. Aisthesis Verlag, Bielefeld.

Selbmann, R. (2018). Das Buddenbrookhaus. In Mattern, N. und Neuhaus, S., Herausgeber, *Buddenbrooks-Handbuch*, Seiten 11–14. J.B. Metzler, Stuttgart.

Sharnagat, R. (2014). Named Entity Recognition: A Literature Survey. http://www.cfilt.iitb.ac.in/resources/surveys/rahul-ner-survey.pdf [15.3.2020].

Simmel, G. (2006). Über räumliche Projektionen sozialer Formen. In Dünne, J. und Günzel, S., Herausgeber, *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften*, Seite 304. Suhrkamp, Frankfurt am Main.

Simmler, S., Thorsten, V., and Pielström, S. (2019). Topic Modeling with Interactive Visualizations in a GUI Tool. DOI https://doi.org/10.34894/ENV3TX.

Simons, O. (2015). Nicht-euklidische Räume. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 272–283. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

Stanzel, F. K. (2008). *Theorie des Erzählens*. UTB: Literaturwissenschaft. Vandenhoeck & Ruprecht, Göttingen, 8. aufl.

Steen, G., editor (2010). *A method for linguistic metaphor identification: from MIP to MIPVU*. Number v. 14 in Converging evidence in language and communication research. John Benjamins Pub. Co, Amsterdam; Philadelphia.

Sutton, C. and Mccallum, A. (2007). An Introduction to Conditional Random Fields for Relational Learning. In *Introduction to Statistical Relational Learning*. http://arxiv.org/abs/1011.4088 [26.2.2020].

Sutton, C. and McCallum, A. (2010). An Introduction to Conditional Random Fields. *arXiv:1011.4088 [stat]*. arXiv: 1011.4088.

Teuber, B. (2015). Rom Imperium Romanum. In Dünne, J. und Mahler, A., Herausgeber, *Handbuch Literatur & Raum*. Seiten 324–334. Handbücher zur kulturwissenschaftlichen Philologie 3. de Gruyter, Berlin; Boston.

TextGrid Repository (2012). Digitale Bibliothek. Textgrid. https://hdl.handle.net/11858/00-1734-0000-0003-84E5-B.

Thaller, M. (2021). Über Metaphern und die Voraussetzungen für ihre Verwendung in der Informationstechnologie. In Flüh, M., Jacke, J., Horstmann, J., and Schumacher, M., Herausgeber, *Toward Undogmatic Reading*, Seiten 81–102. Hamburg University Press, Hamburg.

Todorov, T. (1966). Les catégories du récit littéraire. *Communications*, 8(1):125–151. Publisher: Persée – Portail des revues scientifiques en SHS.

Tuan, Y.-F. (1979). *Space and place the perspective of experience*. Arnold, London, 2nd print edition.

Vala, H., Jurgens, D., Piper, A., and Ruths, D. (2015). Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On The Difficulty of Detecting Characters in Literary Texts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 769–774, Lisbon, Portugal. Association for Computational Linguistics.

Vauth, M., Meister, M., Hatzel, H. O., Gerstorfer, D., and Gius, E. (2022). GitMA. DOI https://doi.org/10.5281/ZENODO.6330464.

Viehhauser, G. (2020). Zur Erkennung von Raum in narrativen Texten: Spatial frames und Raumsemantik als Modelle für eine digitale Narratologie des Raums. In Reiter, N., Pichler, A. und Kuhn, J., Herausgeber, *Reflektierte algorithmische Textanalyse*, Seiten 373–388. De Gruyter, Berlin.

von der Thüsen, J. (1969). ‚Der Romananfang in „Wilhelm Meisters Lehrjahren“‘, Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte, 43(4), Seiten 622–630. DOI https://doi.org/10.1007/BF03376395.

Waerden, B. L. v. d. (1956). *Erwachende Wissenschaft*. Wissenschaft und Kultur. Birkhäuser, Basel.

Wallace, B. C. (2012). Multiple Narrative Disentanglement: Unraveling Infinite Jest. In *NAACL*.

Warhol, R. (2012). Narrative Worlds. In Phelan, J., Herman, D., Rabinowitz, P., Richardson, B., and Warhol, R., editors, *Narrative Theory. Core Concepts and Critical Debates*, pages 92–97. Ohio State University Press, Columbus.

Weber, D. and Pötzl, J. (2014). NERU: Named Entity Recognition for German. https://sites.google.com/site/germeval2014ner/ws-program [14.2.2020].

Wenz, K. (1997). *Raum, Raumsprache und Sprachräume zur Textsemiotik der Raumbeschreibung*. Kodikas, Code / Supplement 22. Narr, Tübingen.

Whybra, D. R. (2005). In Wintgens, H.-H., Herausgeber, Hogwarts – der unberührte Ort. *Literarische Orte – Orte der Literatur*, Seiten 235–247. Universitätsverlag, Hildesheim.

Wiedemann, G., Jindal, R., and Biemann, C. (2018). microNER: A Micro-Service for German Named Entity Recognition based on BiLSTM-CRF. page 7.

Wintgens, H.-H. (2005). Arthur Schnitzler und der Tatort Wien. In Wintgens, H.-H., Herausgeber, *Literarische Orte – Orte der Literatur*, Seiten 185–201. Universitätsverlag, Hildesheim.

Wissler, L., Almashraee, M., Monett, D., and Paschke, A. (2014). *The Gold Standard in Corpus Annotation*. DOI https://doi.org/10.13140/2.1.4316.3523.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zetzmann, V. (2020). Von Argos nach Athen: Von Manipulation zu Polis-Rhetorik zwischen Aischylos' Agamemnon und Eumeniden. In Cecconi, P. und Tornau, Ch., Herausgeber, *Städte und Stadtstaaten zwischen Mythos, Literatur und Propaganda*. Seiten 75–106. De Gruyter, Berlin.

Zoran, G. (1984). Towards a Theory of Space in Narrative. *Poetics Today*, 5(2):309.